

COMPOSITIONS AND METHODS FOR THE THERAPY AND
DIAGNOSIS OF LUNG CANCER

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Patent Application No. 09/849,626, filed May 3, 2001; which is a continuation-in-part of U.S. Patent Application No. 09/736,457, filed December 13, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/702,705, filed October 30, 2000; which is a continuation-
10 in-part of U.S. Patent Application No. 09/677,419, filed October 6, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/671,325, filed September 26, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/658,824, filed September 8, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/651,563, filed August 29, 2000; which is a continuation-in-part of U.S. Patent
15 Application No. 09/614,124, filed July 11, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/589,184, filed June 5, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/560,406, filed April 27, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/546,259, filed April 10, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/533,077, filed March 22, 2000;
20 which is a continuation-in-part of U.S. Patent Application No. 09/519,642, filed March 6, 2000; which is a continuation-in-part of U.S. Patent Application No. 09/476,300, filed December 30, 1999; which is a continuation-in-part of U.S. Patent Application No. 09/466,867, filed December 17, 1999; which is a continuation-in-part of U.S. Patent Application 09/419,356, filed October 15, 1999; which is a continuation-in-part of U.S.
25 Patent Application No. 09/346,492, filed June 30, 1999, all now pending, which applications are incorporated herein by reference in their entirety.

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to therapy and diagnosis of cancer, such as lung cancer. The invention is more specifically related to polypeptides, comprising at least a portion of a lung tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides are useful in pharmaceutical compositions, *e.g.*, vaccines, and other compositions for the diagnosis and treatment of lung cancer.

10 BACKGROUND OF THE INVENTION

Cancer is a significant health problem throughout the world. Although advances have been made in detection and therapy of cancer, no vaccine or other universally successful method for prevention or treatment is currently available. Current therapies, which are generally based on a combination of chemotherapy or surgery and radiation, continue to prove inadequate in many patients.

Lung cancer is the primary cause of cancer death among both men and women in the U.S., with an estimated 172,000 new cases being reported in 1994. The five-year survival rate among all lung cancer patients, regardless of the stage of disease at diagnosis, is only 13%. This contrasts with a five-year survival rate of 46% among cases detected while the disease is still localized. However, only 16% of lung cancers are discovered before the disease has spread.

Early detection is difficult since clinical symptoms are often not seen until the disease has reached an advanced stage. Currently, diagnosis is aided by the use of chest x-rays, analysis of the type of cells contained in sputum and fiberoptic examination of the bronchial passages. Treatment regimens are determined by the type and stage of the cancer, and include surgery, radiation therapy and/or chemotherapy.

In spite of considerable research into therapies for this and other cancers, lung cancer remains difficult to diagnose and treat effectively. Accordingly, there is a need in the art for improved methods for detecting and treating such cancers. The present invention fulfills these needs and further provides other related advantages.

5 SUMMARY OF THE INVENTION

In one aspect, the present invention provides polynucleotide compositions comprising a sequence selected from the group consisting of:

(a) sequences provided in SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002;

(b) complements of the sequences provided in SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002;

(c) sequences consisting of at least 20, 25, 30, 35, 40, 45, 50, 75 and 100 contiguous residues of a sequence provided in SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002;

(d) sequences that hybridize to a sequence provided in SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002, under moderate or highly stringent conditions;

(e) sequences having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to a sequence of SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788,

790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002;

(f) degenerate variants of a sequence provided in SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002.

10 In one preferred embodiment, the polynucleotide compositions of the invention are expressed in at least about 20%, more preferably in at least about 30%, and most preferably in at least about 50% of lung tumors samples tested, at a level that is at least about 2-fold, preferably at least about 5-fold, and most preferably at least about 10-fold higher than that for normal tissues.

15 The present invention, in another aspect, provides polypeptide compositions comprising an amino acid sequence that is encoded by a polynucleotide sequence described above.

The present invention further provides polypeptide compositions comprising an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NOs:324-340, 783, 786, 787, 789, 791, 793, 795, 797-799, 805, 806, 809, 827, 1667, 1670-1675, 1677-1679, 1806-1822, 1825, 1830-1833, 1834-1856, 1863, 1864, 1869-1872, 1874, 1876, 1878, 1880, 1882, 1884-1890, 1901-1909, 1913, 1917, 1921, 1925-1930, 1932, 1934, 1937, 1940 and 1942-1973.

25 In certain preferred embodiments, the polypeptides and/or polynucleotides of the present invention are immunogenic, *i.e.*, they are capable of eliciting an immune response, particularly a humoral and/or cellular immune response, as further described herein.

The present invention further provides fragments, variants and/or derivatives of the disclosed polypeptide and/or polynucleotide sequences. wherein the fragments,

variants and/or derivatives preferably have a level of immunogenic activity of at least about 50%, preferably at least about 70% and more preferably at least about 90% of the level of immunogenic activity of a polypeptide sequence set forth in SEQ ID NOs:324-340, 783, 786, 787, 789, 791, 793, 795, 797-799, 805, 806, 809, 827, 1667, 1670-1675, 1677-1679, 5 1806-1822, 1825, 1830-1833, 1834-1856, 1863, 1864, 1869-1872, 1874, 1876, 1878, 1880, 1882, 1884-1890, 1901-1909, 1913, 1917, 1921, 1925-1930, 1932, 1934, 1937, 1940 and 1942-1973 or a polypeptide sequence encoded by a polynucleotide sequence set forth in SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 10 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002.

The present invention further provides polynucleotides that encode a polypeptide described above, expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

15 Within other aspects, the present invention provides pharmaceutical compositions comprising a polypeptide or polynucleotide as described above and a physiologically acceptable carrier.

Within a related aspect of the present invention, the pharmaceutical compositions, *e.g.*, vaccine compositions, are provided for prophylactic or therapeutic 20 applications. Such compositions generally comprise an immunogenic polypeptide or polynucleotide of the invention and an immunostimulant, such as an adjuvant.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment thereof that specifically binds to a polypeptide of the present invention, or a fragment thereof; and (b) a physiologically 25 acceptable carrier.

Within further aspects, the present invention provides pharmaceutical compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Illustrative

antigen presenting cells include dendritic cells, macrophages, monocytes, fibroblasts and B cells.

Within related aspects, pharmaceutical compositions are provided that comprise: (a) an antigen presenting cell that expresses a polypeptide as described above
 5 and (b) an immunostimulant.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides encoding such fusion proteins, typically in the form of pharmaceutical compositions, *e.g.*, vaccine compositions, comprising a physiologically acceptable carrier and/or an immunostimulant.
 10 The fusions proteins may comprise multiple immunogenic polypeptides or portions/variants thereof, as described herein, and may further comprise one or more polypeptide segments for facilitating the expression, purification and/or immunogenicity of the polypeptide(s).

Within further aspects, the present invention provides methods for stimulating an immune response in a patient, preferably a T cell response in a human
 15 patient, comprising administering a pharmaceutical composition described herein. The patient may be afflicted with lung cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a
 20 pharmaceutical composition as recited above. The patient may be afflicted with lung cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological sample
 25 with T cells that specifically react with a polypeptide of the present invention, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or
 5 expanding T cells specific for a polypeptide of the present invention, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described
 10 above are also provided.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the
 15 development of a cancer in a patient, comprising the steps of: (a) incubating CD4⁺ and/or CD8⁺ T cells isolated from a patient with one or more of: (i) a polypeptide comprising at least an immunogenic portion of polypeptide disclosed herein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated
 20 T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

Within further aspects, the present invention provides methods for determining the presence or absence of a cancer, preferably a lung cancer, in a patient comprising: (a) contacting a biological sample obtained from a patient with a binding agent
 25 that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polypeptide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d)

comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Within further aspects, the present invention provides antibodies, such as monoclonal antibodies, that bind to a polypeptide as described above, as well as diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more oligonucleotide probes or primers as described above are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

10 SEQUENCE IDENTIFIERS

SEQ ID NO:1 is the determined cDNA sequence for clone #19038, also referred to as L845P.

SEQ ID NO:2 is the determined cDNA sequence for clone #19036.

SEQ ID NO:3 is the determined cDNA sequence for clone #19034.

15 SEQ ID NO:4 is the determined cDNA sequence for clone #19033.

SEQ ID NO:5 is the determined cDNA sequence for clone #19032.

SEQ ID NO:6 is the determined cDNA sequence for clone #19030, also referred to as L559S.

SEQ ID NO:7 is the determined cDNA sequence for clone #19029.

20 SEQ ID NO:8 is the determined cDNA sequence for clone #19025.

SEQ ID NO:9 is the determined cDNA sequence for clone #19023.

SEQ ID NO:10 is the determined cDNA sequence for clone #18929.

SEQ ID NO:11 is the determined cDNA sequence for clone #19010.

SEQ ID NO:12 is the determined cDNA sequence for clone #19009.

25 SEQ ID NO:13 is the determined cDNA sequence for clones #19005, 19007, 19016 and 19017.

SEQ ID NO:14 is the determined cDNA sequence for clone #19004.

SEQ ID NO:15 is the determined cDNA sequence for clones #19002 and 18965.

- SEQ ID NO:16 is the determined cDNA sequence for clone #18998.
- SEQ ID NO:17 is the determined cDNA sequence for clone #18997.
- SEQ ID NO:18 is the determined cDNA sequence for clone #18996.
- SEQ ID NO:19 is the determined cDNA sequence for clone #18995.
- 5 SEQ ID NO:20 is the determined cDNA sequence for clone #18994, also known as L846P.
- SEQ ID NO:21 is the determined cDNA sequence for clone #18992.
- SEQ ID NO:22 is the determined cDNA sequence for clone #18991.
- SEQ ID NO:23 is the determined cDNA sequence for clone #18990, also referred to as clone #20111.
- 10 SEQ ID NO:24 is the determined cDNA sequence for clone #18987.
- SEQ ID NO:25 is the determined cDNA sequence for clone #18985, also referred as L839P.
- SEQ ID NO:26 is the determined cDNA sequence for clone #18984, also referred to as L847P.
- 15 SEQ ID NO:27 is the determined cDNA sequence for clone #18983.
- SEQ ID NO:28 is the determined cDNA sequence for clones #18976 and 18980.
- SEQ ID NO:29 is the determined cDNA sequence for clone #18975.
- SEQ ID NO:30 is the determined cDNA sequence for clone #18974.
- SEQ ID NO:31 is the determined cDNA sequence for clone #18973.
- 20 SEQ ID NO:32 is the determined cDNA sequence for clone #18972.
- SEQ ID NO:33 is the determined cDNA sequence for clone #18971, also referred to as L801P.
- SEQ ID NO:34 is the determined cDNA sequence for clone #18970.
- SEQ ID NO:35 is the determined cDNA sequence for clone #18966.
- 25 SEQ ID NO:36 is the determined cDNA sequence for clones #18964, 18968 and 19039.
- SEQ ID NO:37 is the determined cDNA sequence for clone #18960.
- SEQ ID NO:38 is the determined cDNA sequence for clone #18959.
- SEQ ID NO:39 is the determined cDNA sequence for clones #18958 and 18982.
- SEQ ID NO:40 is the determined cDNA sequence for clones #18956 and 19015.

SEQ ID NO:41 is the determined cDNA sequence for clone #18954, also referred to L848P.

SEQ ID NO:42 is the determined cDNA sequence for clone #18951.

SEQ ID NO:43 is the determined cDNA sequence for clone #18950.

- 5 SEQ ID NO:44 is the determined cDNA sequence for clones #18949 and 19024, also referred to as L844P.

SEQ ID NO:45 is the determined cDNA sequence for clone #18948.

SEQ ID NO:46 is the determined cDNA sequence for clone #18947, also referred to as L840P.

- 10 SEQ ID NO:47 is the determined cDNA sequence for clones #18946, 18953, 18969 and 19027.

SEQ ID NO:48 is the determined cDNA sequence for clone #18942.

SEQ ID NO:49 is the determined cDNA sequence for clone #18940, 18962, 18963, 19006, 19008, 19000, and 19031.

- 15 SEQ ID NO:50 is the determined cDNA sequence for clone #18939.

SEQ ID NO:51 is the determined cDNA sequence for clones #18938 and 18952.

SEQ ID NO:52 is the determined cDNA sequence for clone #18938.

SEQ ID NO:53 is the determined cDNA sequence for clone #18937.

- 20 SEQ ID NO:54 is the determined cDNA sequence for clones #18934, 18935, 18993 and 19022, also referred to as L548S.

SEQ ID NO:55 is the determined cDNA sequence for clone #18932.

SEQ ID NO:56 is the determined cDNA sequence for clones #18931 and 18936.

SEQ ID NO:57 is the determined cDNA sequence for clone #18930.

- 25 SEQ ID NO:58 is the determined cDNA sequence for clone #19014 (this sequence has homology to clone L773P, which is also described in co-pending U.S. application 09/285,479, filed April 2, 1999).

SEQ ID NO:59 is the determined cDNA sequence for clone #19127.

SEQ ID NO:60 is the determined cDNA sequence for clones #19057 and 19064.

SEQ ID NO:61 is the determined cDNA sequence for clone #19122.

- SEQ ID NO:62 is the determined cDNA sequence for clones #19120 and 18121.
- SEQ ID NO:63 is the determined cDNA sequence for clone #19118.
- SEQ ID NO:64 is the determined cDNA sequence for clone #19117.
- SEQ ID NO:65 is the determined cDNA sequence for clone #19116.
- 5 SEQ ID NO:66 is the determined cDNA sequence for clone #19114.
- SEQ ID NO:67 is the determined cDNA sequence for clone #19112, also known as L561S.
- SEQ ID NO:68 is the determined cDNA sequence for clone #19110.
- SEQ ID NO:69 is the determined cDNA sequence for clone #19107, also referred to as L552S.
- 10 SEQ ID NO:70 is the determined cDNA sequence for clone #19106, also referred to as L547S.
- SEQ ID NO:71 is the determined cDNA sequence for clones #19105 and 19111.
- SEQ ID NO:72 is the determined cDNA sequence for clone #19099.
- SEQ ID NO:73 is the determined cDNA sequence for clones #19095, 19104 and 19125,
- 15 also referred to as L549S.
- SEQ ID NO:74 is the determined cDNA sequence for clone #19094.
- SEQ ID NO:75 is the determined cDNA sequence for clones #19089 and 19101.
- SEQ ID NO:76 is the determined cDNA sequence for clone #19088.
- SEQ ID NO:77 is the determined cDNA sequence for clones #19087, 19092, 19096, 19100
- 20 and 19119.
- SEQ ID NO:78 is the determined cDNA sequence for clone #19086.
- SEQ ID NO:79 is the determined cDNA sequence for clone #19085, also referred to as L550S.
- SEQ ID NO:80 is the determined cDNA sequence for clone #19084, also referred to as
- 25 clone #19079.
- SEQ ID NO:81 is the determined cDNA sequence for clone #19082.
- SEQ ID NO:82 is the determined cDNA sequence for clone #19080.
- SEQ ID NO:83 is the determined cDNA sequence for clone #19077.

SEQ ID NO:84 is the determined cDNA sequence for clone #19076, also referred to as L551S.

SEQ ID NO:85 is the determined cDNA sequence for clone #19074, also referred to as clone #20102.

- 5 SEQ ID NO:86 is the determined cDNA sequence for clone #19073, also referred to as L560S.

SEQ ID NO:87 is the determined cDNA sequence for clones #19072 and 19115.

SEQ ID NO:88 is the determined cDNA sequence for clone #19071.

SEQ ID NO:89 is the determined cDNA sequence for clone #19070.

- 10 SEQ ID NO:90 is the determined cDNA sequence for clone #19069.

SEQ ID NO:91 is the determined cDNA sequence for clone #19068, also referred to L563S.

SEQ ID NO:92 is the determined cDNA sequence for clone #19066.

SEQ ID NO:93 is the determined cDNA sequence for clone #19065.

- 15 SEQ ID NO:94 is the determined cDNA sequence for clone #19063.

SEQ ID NO:95 is the determined cDNA sequence for clones #19061, 19081, 19108 and 19109.

SEQ ID NO:96 is the determined cDNA sequence for clones #19060, 19067 and 19083, also referred to as L548S.

- 20 SEQ ID NO:97 is the determined cDNA sequence for clones #19059 and 19062.

SEQ ID NO:98 is the determined cDNA sequence for clone #19058.

SEQ ID NO:99 is the determined cDNA sequence for clone #19124.

SEQ ID NO:100 is the determined cDNA sequence for clone #18929.

SEQ ID NO:101 is the determined cDNA sequence for clone #18422.

- 25 SEQ ID NO:102 is the determined cDNA sequence for clone #18425.

SEQ ID NO:103 is the determined cDNA sequence for clone #18431.

SEQ ID NO:104 is the determined cDNA sequence for clone #18433.

SEQ ID NO:105 is the determined cDNA sequence for clone #18444.

SEQ ID NO:106 is the determined cDNA sequence for clone #18449.

- SEQ ID NO:107 is the determined cDNA sequence for clone #18451.
 SEQ ID NO:108 is the determined cDNA sequence for clone #18452.
 SEQ ID NO:109 is the determined cDNA sequence for clone #18455.
 SEQ ID NO:110 is the determined cDNA sequence for clone #18457.
 5 SEQ ID NO:111 is the determined cDNA sequence for clone #18466.
 SEQ ID NO:112 is the determined cDNA sequence for clone #18468.
 SEQ ID NO:113 is the determined cDNA sequence for clone #18471.
 SEQ ID NO:114 is the determined cDNA sequence for clone #18475.
 SEQ ID NO:115 is the determined cDNA sequence for clone #18476.
 10 SEQ ID NO:116 is the determined cDNA sequence for clone #18477.
 SEQ ID NO:117 is the determined cDNA sequence for clone #20631.
 SEQ ID NO:118 is the determined cDNA sequence for clone #20634.
 SEQ ID NO:119 is the determined cDNA sequence for clone #20635.
 SEQ ID NO:120 is the determined cDNA sequence for clone #20637.
 15 SEQ ID NO:121 is the determined cDNA sequence for clone #20638.
 SEQ ID NO:122 is the determined cDNA sequence for clone #20643.
 SEQ ID NO:123 is the determined cDNA sequence for clone #20652.
 SEQ ID NO:124 is the determined cDNA sequence for clone #20653.
 SEQ ID NO:125 is the determined cDNA sequence for clone #20657.
 20 SEQ ID NO:126 is the determined cDNA sequence for clone #20658.
 SEQ ID NO:127 is the determined cDNA sequence for clone #20660.
 SEQ ID NO:128 is the determined cDNA sequence for clone #20661.
 SEQ ID NO:129 is the determined cDNA sequence for clone #20663.
 SEQ ID NO:130 is the determined cDNA sequence for clone #20665.
 25 SEQ ID NO:131 is the determined cDNA sequence for clone #20670.
 SEQ ID NO:132 is the determined cDNA sequence for clone #20671.
 SEQ ID NO:133 is the determined cDNA sequence for clone #20672.
 SEQ ID NO:134 is the determined cDNA sequence for clone #20675.
 SEQ ID NO:135 is the determined cDNA sequence for clone #20679.

- SEQ ID NO:136 is the determined cDNA sequence for clone #20681.
 SEQ ID NO:137 is the determined cDNA sequence for clone #20682.
 SEQ ID NO:138 is the determined cDNA sequence for clone #20684.
 SEQ ID NO:139 is the determined cDNA sequence for clone #20685.
 5 SEQ ID NO:140 is the determined cDNA sequence for clone #20689.
 SEQ ID NO:141 is the determined cDNA sequence for clone #20699.
 SEQ ID NO:142 is the determined cDNA sequence for clone #20701.
 SEQ ID NO:143 is the determined cDNA sequence for clone #20702.
 SEQ ID NO:144 is the determined cDNA sequence for clone #20708.
 10 SEQ ID NO:145 is the determined cDNA sequence for clone #20715.
 SEQ ID NO:146 is the determined cDNA sequence for clone #20716.
 SEQ ID NO:147 is the determined cDNA sequence for clone #20719.
 SEQ ID NO:148 is the determined cDNA sequence for clone #19129.
 SEQ ID NO:149 is the determined cDNA sequence for clone #19131.1.
 15 SEQ ID NO:150 is the determined cDNA sequence for clone #19132.2.
 SEQ ID NO:151 is the determined cDNA sequence for clone #19133.
 SEQ ID NO:152 is the determined cDNA sequence for clone #19134.2.
 SEQ ID NO:153 is the determined cDNA sequence for clone #19135.2.
 SEQ ID NO:154 is the determined cDNA sequence for clone #19137.
 20 SEQ ID NO:155 is a first determined cDNA sequence for clone #19138.1.
 SEQ ID NO:156 is a second determined cDNA sequence for clone #19138.2.
 SEQ ID NO:157 is the determined cDNA sequence for clone #19139.
 SEQ ID NO:158 is a first determined cDNA sequence for clone #19140.1.
 SEQ ID NO:159 is a second determined cDNA sequence for clone #19140.2.
 25 SEQ ID NO:160 is the determined cDNA sequence for clone #19141.
 SEQ ID NO:161 is the determined cDNA sequence for clone #19143.
 SEQ ID NO:162 is the determined cDNA sequence for clone #19144.
 SEQ ID NO:163 is a first determined cDNA sequence for clone #19145.1.
 SEQ ID NO:164 is a second determined cDNA sequence for clone #19145.2.

- SEQ ID NO:165 is the determined cDNA sequence for clone #19146.
- SEQ ID NO:166 is the determined cDNA sequence for clone #19149.1.
- SEQ ID NO:167 is the determined cDNA sequence for clone #19152.
- SEQ ID NO:168 is a first determined cDNA sequence for clone #19153.1.
- 5 SEQ ID NO:169 is a second determined cDNA sequence for clone #19153.2.
- SEQ ID NO:170 is the determined cDNA sequence for clone #19155.
- SEQ ID NO:171 is the determined cDNA sequence for clone #19157.
- SEQ ID NO:172 is the determined cDNA sequence for clone #19159.
- SEQ ID NO:173 is the determined cDNA sequence for clone #19160.
- 10 SEQ ID NO:174 is a first determined cDNA sequence for clone #19161.1.
- SEQ ID NO:175 is a second determined cDNA sequence for clone #19161.2.
- SEQ ID NO:176 is the determined cDNA sequence for clone #19162.1.
- SEQ ID NO:177 is the determined cDNA sequence for clone #19166.
- SEQ ID NO:178 is the determined cDNA sequence for clone #19169.
- 15 SEQ ID NO:179 is the determined cDNA sequence for clone #19171.
- SEQ ID NO:180 is a first determined cDNA sequence for clone #19173.1.
- SEQ ID NO:181 is a second determined cDNA sequence for clone #19173.2.
- SEQ ID NO:182 is the determined cDNA sequence for clone #19174.1.
- SEQ ID NO:183 is the determined cDNA sequence for clone #19175.
- 20 SEQ ID NO:184 is the determined cDNA sequence for clone #19177.
- SEQ ID NO:185 is the determined cDNA sequence for clone #19178.
- SEQ ID NO:186 is the determined cDNA sequence for clone #19179.1.
- SEQ ID NO:187 is the determined cDNA sequence for clone #19179.2.
- SEQ ID NO:188 is the determined cDNA sequence for clone #19180.
- 25 SEQ ID NO:189 is a first determined cDNA sequence for clone #19182.1.
- SEQ ID NO:190 is a second determined cDNA sequence for clone #19182.2.
- SEQ ID NO:191 is the determined cDNA sequence for clone #19183.1.
- SEQ ID NO:192 is the determined cDNA sequence for clone #19185.1.
- SEQ ID NO:193 is the determined cDNA sequence for clone #19187.

- SEQ ID NO:194 is the determined cDNA sequence for clone #19188.
- SEQ ID NO:195 is the determined cDNA sequence for clone #19190.
- SEQ ID NO:196 is the determined cDNA sequence for clone #19191.
- SEQ ID NO:197 is the determined cDNA sequence for clone #19192.
- 5 SEQ ID NO:198 is the determined cDNA sequence for clone #19193.
- SEQ ID NO:199 is a first determined cDNA sequence for clone #19194.1.
- SEQ ID NO:200 is a second determined cDNA sequence for clone #19194.2.
- SEQ ID NO:201 is the determined cDNA sequence for clone #19197.
- SEQ ID NO:202 is a first determined cDNA sequence for clone #19200.1.
- 10 SEQ ID NO:203 is a second determined cDNA sequence for clone #19200.2.
- SEQ ID NO:204 is the determined cDNA sequence for clone #19202.
- SEQ ID NO:205 is a first determined cDNA sequence for clone #19204.1.
- SEQ ID NO:206 is a second determined cDNA sequence for clone #19204.2.
- SEQ ID NO:207 is the determined cDNA sequence for clone #19205.
- 15 SEQ ID NO:208 is a first determined cDNA sequence for clone #19206.1.
- SEQ ID NO:209 is a second determined cDNA sequence for clone #19206.2.
- SEQ ID NO:210 is the determined cDNA sequence for clone #19207.
- SEQ ID NO:211 is the determined cDNA sequence for clone #19208.
- SEQ ID NO:212 is a first determined cDNA sequence for clone #19211.1.
- 20 SEQ ID NO:213 is a second determined cDNA sequence for clone #19211.2.
- SEQ ID NO:214 is a first determined cDNA sequence for clone #19214.1.
- SEQ ID NO:215 is a second determined cDNA sequence for clone #19214.2.
- SEQ ID NO:216 is the determined cDNA sequence for clone #19215.
- SEQ ID NO:217 is a first determined cDNA sequence for clone #19217. 2.
- 25 SEQ ID NO:218 is a second determined cDNA sequence for clone #19217.2.
- SEQ ID NO:219 is a first determined cDNA sequence for clone #19218.1.
- SEQ ID NO:220 is a second determined cDNA sequence for clone #19218.2.
- SEQ ID NO:221 is a first determined cDNA sequence for clone #19220.1.
- SEQ ID NO:222 is a second determined cDNA sequence for clone #19220.2.

- SEQ ID NO:223 is the determined cDNA sequence for clone #22015.
 SEQ ID NO:224 is the determined cDNA sequence for clone #22017.
 SEQ ID NO:225 is the determined cDNA sequence for clone #22019.
 SEQ ID NO:226 is the determined cDNA sequence for clone #22020.
 5 SEQ ID NO:227 is the determined cDNA sequence for clone #22023.
 SEQ ID NO:228 is the determined cDNA sequence for clone #22026.
 SEQ ID NO:229 is the determined cDNA sequence for clone #22027.
 SEQ ID NO:230 is the determined cDNA sequence for clone #22028.
 SEQ ID NO:231 is the determined cDNA sequence for clone #22032.
 10 SEQ ID NO:232 is the determined cDNA sequence for clone #22037.
 SEQ ID NO:233 is the determined cDNA sequence for clone #22045.
 SEQ ID NO:234 is the determined cDNA sequence for clone #22048.
 SEQ ID NO:235 is the determined cDNA sequence for clone #22050.
 SEQ ID NO:236 is the determined cDNA sequence for clone #22052.
 15 SEQ ID NO:237 is the determined cDNA sequence for clone #22053.
 SEQ ID NO:238 is the determined cDNA sequence for clone #22057.
 SEQ ID NO:239 is the determined cDNA sequence for clone #22066.
 SEQ ID NO:240 is the determined cDNA sequence for clone #22077.
 SEQ ID NO:241 is the determined cDNA sequence for clone #22085.
 20 SEQ ID NO:242 is the determined cDNA sequence for clone #22105.
 SEQ ID NO:243 is the determined cDNA sequence for clone #22108.
 SEQ ID NO:244 is the determined cDNA sequence for clone #22109.
 SEQ ID NO:245 is the determined cDNA sequence for clone #24842.
 SEQ ID NO:246 is the determined cDNA sequence for clone #24843.
 25 SEQ ID NO:247 is the determined cDNA sequence for clone #24845.
 SEQ ID NO:248 is the determined cDNA sequence for clone #24851.
 SEQ ID NO:249 is the determined cDNA sequence for clone #24852.
 SEQ ID NO:250 is the determined cDNA sequence for clone #24853.
 SEQ ID NO:251 is the determined cDNA sequence for clone #24854.

- SEQ ID NO:252 is the determined cDNA sequence for clone #24855.
 SEQ ID NO:253 is the determined cDNA sequence for clone #24860.
 SEQ ID NO:254 is the determined cDNA sequence for clone #24864.
 SEQ ID NO:255 is the determined cDNA sequence for clone #24866.
 5 SEQ ID NO:256 is the determined cDNA sequence for clone #24867.
 SEQ ID NO:257 is the determined cDNA sequence for clone #24868.
 SEQ ID NO:258 is the determined cDNA sequence for clone #24869.
 SEQ ID NO:259 is the determined cDNA sequence for clone #24870.
 SEQ ID NO:260 is the determined cDNA sequence for clone #24872.
 10 SEQ ID NO:261 is the determined cDNA sequence for clone #24873.
 SEQ ID NO:262 is the determined cDNA sequence for clone #24875.
 SEQ ID NO:263 is the determined cDNA sequence for clone #24882.
 SEQ ID NO:264 is the determined cDNA sequence for clone #24885.
 SEQ ID NO:265 is the determined cDNA sequence for clone #24886.
 15 SEQ ID NO:266 is the determined cDNA sequence for clone #24887.
 SEQ ID NO:267 is the determined cDNA sequence for clone #24888.
 SEQ ID NO:268 is the determined cDNA sequence for clone #24890.
 SEQ ID NO:269 is the determined cDNA sequence for clone #24896.
 SEQ ID NO:270 is the determined cDNA sequence for clone #24897.
 20 SEQ ID NO:271 is the determined cDNA sequence for clone #24899.
 SEQ ID NO:272 is the determined cDNA sequence for clone #24901.
 SEQ ID NO:273 is the determined cDNA sequence for clone #24902.
 SEQ ID NO:274 is the determined cDNA sequence for clone #24906.
 SEQ ID NO:275 is the determined cDNA sequence for clone #24912.
 25 SEQ ID NO:276 is the determined cDNA sequence for clone #24913.
 SEQ ID NO:277 is the determined cDNA sequence for clone #24920.
 SEQ ID NO:278 is the determined cDNA sequence for clone #24927.
 SEQ ID NO:279 is the determined cDNA sequence for clone #24930.
 SEQ ID NO:280 is the determined cDNA sequence for clone #26938.

- SEQ ID NO:281 is the determined cDNA sequence for clone #26939.
 SEQ ID NO:282 is the determined cDNA sequence for clone #26943.
 SEQ ID NO:283 is the determined cDNA sequence for clone #26948.
 SEQ ID NO:284 is the determined cDNA sequence for clone #26951.
 5 SEQ ID NO:285 is the determined cDNA sequence for clone #26955.
 SEQ ID NO:286 is the determined cDNA sequence for clone #26956.
 SEQ ID NO:287 is the determined cDNA sequence for clone #26959.
 SEQ ID NO:288 is the determined cDNA sequence for clone #26961.
 SEQ ID NO:289 is the determined cDNA sequence for clone #26962.
 10 SEQ ID NO:290 is the determined cDNA sequence for clone #26964.
 SEQ ID NO:291 is the determined cDNA sequence for clone #26966.
 SEQ ID NO:292 is the determined cDNA sequence for clone #26968.
 SEQ ID NO:293 is the determined cDNA sequence for clone #26972.
 SEQ ID NO:294 is the determined cDNA sequence for clone #26973.
 15 SEQ ID NO:295 is the determined cDNA sequence for clone #26974.
 SEQ ID NO:296 is the determined cDNA sequence for clone #26976.
 SEQ ID NO:297 is the determined cDNA sequence for clone #26977.
 SEQ ID NO:298 is the determined cDNA sequence for clone #26979.
 SEQ ID NO:299 is the determined cDNA sequence for clone #26980.
 20 SEQ ID NO:300 is the determined cDNA sequence for clone #26981.
 SEQ ID NO:301 is the determined cDNA sequence for clone #26984.
 SEQ ID NO:302 is the determined cDNA sequence for clone #26985.
 SEQ ID NO:303 is the determined cDNA sequence for clone #26986.
 SEQ ID NO:304 is the determined cDNA sequence for clone #26993.
 25 SEQ ID NO:305 is the determined cDNA sequence for clone #26994.
 SEQ ID NO:306 is the determined cDNA sequence for clone #26995.
 SEQ ID NO:307 is the determined cDNA sequence for clone #27003.
 SEQ ID NO:308 is the determined cDNA sequence for clone #27005.
 SEQ ID NO:309 is the determined cDNA sequence for clone #27010.

- SEQ ID NO:310 is the determined cDNA sequence for clone #27011.
 SEQ ID NO:311 is the determined cDNA sequence for clone #27013.
 SEQ ID NO:312 is the determined cDNA sequence for clone #27016
 SEQ ID NO:313 is the determined cDNA sequence for clone #27017.
 5 SEQ ID NO:314 is the determined cDNA sequence for clone #27019.
 SEQ ID NO:315 is the determined cDNA sequence for clone #27028.
 SEQ ID NO:316 is the full-length cDNA sequence for clone #19060.
 SEQ ID NO:317 is the full-length cDNA sequence for clone #18964.
 SEQ ID NO:318 is the full-length cDNA sequence for clone #18929.
 10 SEQ ID NO:319 is the full-length cDNA sequence for clone #18991.
 SEQ ID NO:320 is the full-length cDNA sequence for clone #18996.
 SEQ ID NO:321 is the full-length cDNA sequence for clone #18966.
 SEQ ID NO:322 is the full-length cDNA sequence for clone #18951.
 SEQ ID NO:323 is the full-length cDNA sequence for clone #18973 (also known as
 15 L516S).
 SEQ ID NO:324 is the amino acid sequence for clone #19060.
 SEQ ID NO:325 is the amino acid sequence for clone #19063.
 SEQ ID NO:326 is the amino acid sequence for clone #19077.
 SEQ ID NO:327 is the amino acid sequence for clone #19110.
 20 SEQ ID NO:328 is the amino acid sequence for clone #19122.
 SEQ ID NO:329 is the amino acid sequence for clone #19118.
 SEQ ID NO:330 is the amino acid sequence for clone #19080.
 SEQ ID NO:331 is the amino acid sequence for clone #19127.
 SEQ ID NO:332 is the amino acid sequence for clone #19117.
 25 SEQ ID NO:333 is the amino acid sequence for clone #19095, also referred to L549S.
 SEQ ID NO:334 is the amino acid sequence for clone #18964.
 SEQ ID NO:335 is the amino acid sequence for clone #18929.
 SEQ ID NO:336 is the amino acid sequence for clone #18991.
 SEQ ID NO:337 is the amino acid sequence for clone #18996.

SEQ ID NO:338 is the amino acid sequence for clone #18966.

SEQ ID NO:339 is the amino acid sequence for clone #18951.

SEQ ID NO:340 is the amino acid sequence for clone #18973.

SEQ ID NO:341 is the determined cDNA sequence for clone 26461.

5 SEQ ID NO:342 is the determined cDNA sequence for clone 26462.

SEQ ID NO:343 is the determined cDNA sequence for clone 26463.

SEQ ID NO:344 is the determined cDNA sequence for clone 26464.

SEQ ID NO:345 is the determined cDNA sequence for clone 26465.

SEQ ID NO:346 is the determined cDNA sequence for clone 26466.

10 SEQ ID NO:347 is the determined cDNA sequence for clone 26467.

SEQ ID NO:348 is the determined cDNA sequence for clone 26468.

SEQ ID NO:349 is the determined cDNA sequence for clone 26469.

SEQ ID NO:350 is the determined cDNA sequence for clone 26470.

SEQ ID NO:351 is the determined cDNA sequence for clone 26471.

15 SEQ ID NO:352 is the determined cDNA sequence for clone 26472.

SEQ ID NO:353 is the determined cDNA sequence for clone 26474.

SEQ ID NO:354 is the determined cDNA sequence for clone 26475.

SEQ ID NO:355 is the determined cDNA sequence for clone 26476.

SEQ ID NO:356 is the determined cDNA sequence for clone 26477.

20 SEQ ID NO:357 is the determined cDNA sequence for clone 26478.

SEQ ID NO:358 is the determined cDNA sequence for clone 26479.

SEQ ID NO:359 is the determined cDNA sequence for clone 26480.

SEQ ID NO:360 is the determined cDNA sequence for clone 26481.

SEQ ID NO:361 is the determined cDNA sequence for clone 26482.

25 SEQ ID NO:362 is the determined cDNA sequence for clone 26483.

SEQ ID NO:363 is the determined cDNA sequence for clone 26484.

SEQ ID NO:364 is the determined cDNA sequence for clone 26485.

SEQ ID NO:365 is the determined cDNA sequence for clone 26486.

SEQ ID NO:366 is the determined cDNA sequence for clone 26487.

- SEQ ID NO:367 is the determined cDNA sequence for clone 26488.
 SEQ ID NO:368 is the determined cDNA sequence for clone 26489.
 SEQ ID NO:369 is the determined cDNA sequence for clone 26490.
 SEQ ID NO:370 is the determined cDNA sequence for clone 26491.
 5 SEQ ID NO:371 is the determined cDNA sequence for clone 26492.
 SEQ ID NO:372 is the determined cDNA sequence for clone 26493.
 SEQ ID NO:373 is the determined cDNA sequence for clone 26494.
 SEQ ID NO:374 is the determined cDNA sequence for clone 26495.
 SEQ ID NO:375 is the determined cDNA sequence for clone 26496.
 10 SEQ ID NO:376 is the determined cDNA sequence for clone 26497.
 SEQ ID NO:377 is the determined cDNA sequence for clone 26498.
 SEQ ID NO:378 is the determined cDNA sequence for clone 26499.
 SEQ ID NO:379 is the determined cDNA sequence for clone 26500.
 SEQ ID NO:380 is the determined cDNA sequence for clone 26501.
 15 SEQ ID NO:381 is the determined cDNA sequence for clone 26502.
 SEQ ID NO:382 is the determined cDNA sequence for clone 26503.
 SEQ ID NO:383 is the determined cDNA sequence for clone 26504.
 SEQ ID NO:384 is the determined cDNA sequence for clone 26505.
 SEQ ID NO:385 is the determined cDNA sequence for clone 26506.
 20 SEQ ID NO:386 is the determined cDNA sequence for clone 26507.
 SEQ ID NO:387 is the determined cDNA sequence for clone 26508.
 SEQ ID NO:388 is the determined cDNA sequence for clone 26509.
 SEQ ID NO:389 is the determined cDNA sequence for clone 26511.
 SEQ ID NO:390 is the determined cDNA sequence for clone 26513.
 25 SEQ ID NO:391 is the determined cDNA sequence for clone 26514.
 SEQ ID NO:392 is the determined cDNA sequence for clone 26515.
 SEQ ID NO:393 is the determined cDNA sequence for clone 26516.
 SEQ ID NO:394 is the determined cDNA sequence for clone 26517.
 SEQ ID NO:395 is the determined cDNA sequence for clone 26518.

- SEQ ID NO:396 is the determined cDNA sequence for clone 26519.
 SEQ ID NO:397 is the determined cDNA sequence for clone 26520.
 SEQ ID NO:398 is the determined cDNA sequence for clone 26521.
 SEQ ID NO:399 is the determined cDNA sequence for clone 26522.
 5 SEQ ID NO:400 is the determined cDNA sequence for clone 26523.
 SEQ ID NO:401 is the determined cDNA sequence for clone 26524.
 SEQ ID NO:402 is the determined cDNA sequence for clone 26526.
 SEQ ID NO:403 is the determined cDNA sequence for clone 26527.
 SEQ ID NO:404 is the determined cDNA sequence for clone 26528.
 10 SEQ ID NO:405 is the determined cDNA sequence for clone 26529.
 SEQ ID NO:406 is the determined cDNA sequence for clone 26530.
 SEQ ID NO:407 is the determined cDNA sequence for clone 26532.
 SEQ ID NO:408 is the determined cDNA sequence for clone 26533.
 SEQ ID NO:409 is the determined cDNA sequence for clone 26534.
 15 SEQ ID NO:410 is the determined cDNA sequence for clone 26535.
 SEQ ID NO:411 is the determined cDNA sequence for clone 26536.
 SEQ ID NO:412 is the determined cDNA sequence for clone 26537.
 SEQ ID NO:413 is the determined cDNA sequence for clone 26538.
 SEQ ID NO:414 is the determined cDNA sequence for clone 26540.
 20 SEQ ID NO:415 is the determined cDNA sequence for clone 26541.
 SEQ ID NO:416 is the determined cDNA sequence for clone 26542.
 SEQ ID NO:417 is the determined cDNA sequence for clone 26543.
 SEQ ID NO:418 is the determined cDNA sequence for clone 26544.
 SEQ ID NO:419 is the determined cDNA sequence for clone 26546.
 25 SEQ ID NO:420 is the determined cDNA sequence for clone 26547.
 SEQ ID NO:421 is the determined cDNA sequence for clone 26548.
 SEQ ID NO:422 is the determined cDNA sequence for clone 26549.
 SEQ ID NO:423 is the determined cDNA sequence for clone 26550.
 SEQ ID NO:424 is the determined cDNA sequence for clone 26551.

- SEQ ID NO:425 is the determined cDNA sequence for clone 26552.
 SEQ ID NO:426 is the determined cDNA sequence for clone 26553.
 SEQ ID NO:427 is the determined cDNA sequence for clone 26554.
 SEQ ID NO:428 is the determined cDNA sequence for clone 26556.
- 5 SEQ ID NO:429 is the determined cDNA sequence for clone 26557.
 SEQ ID NO:430 is the determined cDNA sequence for clone 27631.
 SEQ ID NO:431 is the determined cDNA sequence for clone 27632.
 SEQ ID NO:432 is the determined cDNA sequence for clone 27633.
 SEQ ID NO:433 is the determined cDNA sequence for clone 27635.
- 10 SEQ ID NO:434 is the determined cDNA sequence for clone 27636.
 SEQ ID NO:435 is the determined cDNA sequence for clone 27637.
 SEQ ID NO:436 is the determined cDNA sequence for clone 27638.
 SEQ ID NO:437 is the determined cDNA sequence for clone 27639.
 SEQ ID NO:438 is the determined cDNA sequence for clone 27640.
- 15 SEQ ID NO:439 is the determined cDNA sequence for clone 27641.
 SEQ ID NO:440 is the determined cDNA sequence for clone 27642.
 SEQ ID NO:441 is the determined cDNA sequence for clone 27644.
 SEQ ID NO:442 is the determined cDNA sequence for clone 27646.
 SEQ ID NO:443 is the determined cDNA sequence for clone 27647.
- 20 SEQ ID NO:444 is the determined cDNA sequence for clone 27649.
 SEQ ID NO:445 is the determined cDNA sequence for clone 27650.
 SEQ ID NO:446 is the determined cDNA sequence for clone 27651.
 SEQ ID NO:447 is the determined cDNA sequence for clone 27652.
 SEQ ID NO:448 is the determined cDNA sequence for clone 27654.
- 25 SEQ ID NO:449 is the determined cDNA sequence for clone 27655.
 SEQ ID NO:450 is the determined cDNA sequence for clone 27657.
 SEQ ID NO:451 is the determined cDNA sequence for clone 27659.
 SEQ ID NO:452 is the determined cDNA sequence for clone 27665.
 SEQ ID NO:453 is the determined cDNA sequence for clone 27666.

- SEQ ID NO:454 is the determined cDNA sequence for clone 27668.
 SEQ ID NO:455 is the determined cDNA sequence for clone 27670.
 SEQ ID NO:456 is the determined cDNA sequence for clone 27671.
 SEQ ID NO:457 is the determined cDNA sequence for clone 27672.
 5 SEQ ID NO:458 is the determined cDNA sequence for clone 27674.
 SEQ ID NO:459 is the determined cDNA sequence for clone 27677.
 SEQ ID NO:460 is the determined cDNA sequence for clone 27681.
 SEQ ID NO:461 is the determined cDNA sequence for clone 27682.
 SEQ ID NO:462 is the determined cDNA sequence for clone 27683.
 10 SEQ ID NO:463 is the determined cDNA sequence for clone 27686.
 SEQ ID NO:464 is the determined cDNA sequence for clone 27688.
 SEQ ID NO:465 is the determined cDNA sequence for clone 27689.
 SEQ ID NO:466 is the determined cDNA sequence for clone 27690.
 SEQ ID NO:467 is the determined cDNA sequence for clone 27693.
 15 SEQ ID NO:468 is the determined cDNA sequence for clone 27699.
 SEQ ID NO:469 is the determined cDNA sequence for clone 27700.
 SEQ ID NO:470 is the determined cDNA sequence for clone 27702.
 SEQ ID NO:471 is the determined cDNA sequence for clone 27705.
 SEQ ID NO:472 is the determined cDNA sequence for clone 27706.
 20 SEQ ID NO:473 is the determined cDNA sequence for clone 27707.
 SEQ ID NO:474 is the determined cDNA sequence for clone 27708.
 SEQ ID NO:475 is the determined cDNA sequence for clone 27709.
 SEQ ID NO:476 is the determined cDNA sequence for clone 27710.
 SEQ ID NO:477 is the determined cDNA sequence for clone 27711.
 25 SEQ ID NO:478 is the determined cDNA sequence for clone 27712.
 SEQ ID NO:479 is the determined cDNA sequence for clone 27713.
 SEQ ID NO:480 is the determined cDNA sequence for clone 27714.
 SEQ ID NO:481 is the determined cDNA sequence for clone 27715.
 SEQ ID NO:482 is the determined cDNA sequence for clone 27716.

- SEQ ID NO:483 is the determined cDNA sequence for clone 27717.
 SEQ ID NO:484 is the determined cDNA sequence for clone 27718.
 SEQ ID NO:485 is the determined cDNA sequence for clone 27719.
 SEQ ID NO:486 is the determined cDNA sequence for clone 27720.
 5 SEQ ID NO:487 is the determined cDNA sequence for clone 27722.
 SEQ ID NO:488 is the determined cDNA sequence for clone 27723.
 SEQ ID NO:489 is the determined cDNA sequence for clone 27724.
 SEQ ID NO:490 is the determined cDNA sequence for clone 27726.
 SEQ ID NO:491 is the determined cDNA sequence for clone 25015.
 10 SEQ ID NO:492 is the determined cDNA sequence for clone 25016.
 SEQ ID NO:493 is the determined cDNA sequence for clone 25017.
 SEQ ID NO:494 is the determined cDNA sequence for clone 25018
 SEQ ID NO:495 is the determined cDNA sequence for clone 25030.
 SEQ ID NO:496 is the determined cDNA sequence for clone 25033.
 15 SEQ ID NO:497 is the determined cDNA sequence for clone 25034.
 SEQ ID NO:498 is the determined cDNA sequence for clone 25035.
 SEQ ID NO:499 is the determined cDNA sequence for clone 25036.
 SEQ ID NO:500 is the determined cDNA sequence for clone 25037.
 SEQ ID NO:501 is the determined cDNA sequence for clone 25038.
 20 SEQ ID NO:502 is the determined cDNA sequence for clone 25039.
 SEQ ID NO:503 is the determined cDNA sequence for clone 25040.
 SEQ ID NO:504 is the determined cDNA sequence for clone 25042.
 SEQ ID NO:505 is the determined cDNA sequence for clone 25043.
 SEQ ID NO:506 is the determined cDNA sequence for clone 25044.
 25 SEQ ID NO:507 is the determined cDNA sequence for clone 25045.
 SEQ ID NO:508 is the determined cDNA sequence for clone 25047.
 SEQ ID NO:509 is the determined cDNA sequence for clone 25048.
 SEQ ID NO:510 is the determined cDNA sequence for clone 25049.
 SEQ ID NO:511 is the determined cDNA sequence for clone 25185.

- SEQ ID NO:512 is the determined cDNA sequence for clone 25186.
 SEQ ID NO:513 is the determined cDNA sequence for clone 25187.
 SEQ ID NO:514 is the determined cDNA sequence for clone 25188.
 SEQ ID NO:515 is the determined cDNA sequence for clone 25189.
 5 SEQ ID NO:516 is the determined cDNA sequence for clone 25190.
 SEQ ID NO:517 is the determined cDNA sequence for clone 25193.
 SEQ ID NO:518 is the determined cDNA sequence for clone 25194.
 SEQ ID NO:519 is the determined cDNA sequence for clone 25196.
 SEQ ID NO:520 is the determined cDNA sequence for clone 25198.
 10 SEQ ID NO:521 is the determined cDNA sequence for clone 25199.
 SEQ ID NO:522 is the determined cDNA sequence for clone 25200.
 SEQ ID NO:523 is the determined cDNA sequence for clone 25202.
 SEQ ID NO:524 is the determined cDNA sequence for clone 25364.
 SEQ ID NO:525 is the determined cDNA sequence for clone 25366.
 15 SEQ ID NO:526 is the determined cDNA sequence for clone 25367.
 SEQ ID NO:527 is the determined cDNA sequence for clone 25368.
 SEQ ID NO:528 is the determined cDNA sequence for clone 25369.
 SEQ ID NO:529 is the determined cDNA sequence for clone 25370.
 SEQ ID NO:530 is the determined cDNA sequence for clone 25371.
 20 SEQ ID NO:531 is the determined cDNA sequence for clone 25372.
 SEQ ID NO:532 is the determined cDNA sequence for clone 25373.
 SEQ ID NO:533 is the determined cDNA sequence for clone 25374.
 SEQ ID NO:534 is the determined cDNA sequence for clone 25376.
 SEQ ID NO:535 is the determined cDNA sequence for clone 25377.
 25 SEQ ID NO:536 is the determined cDNA sequence for clone 25378.
 SEQ ID NO:537 is the determined cDNA sequence for clone 25379.
 SEQ ID NO:538 is the determined cDNA sequence for clone 25380.
 SEQ ID NO:539 is the determined cDNA sequence for clone 25381.
 SEQ ID NO:540 is the determined cDNA sequence for clone 25382.

- SEQ ID NO:541 is the determined cDNA sequence for clone 25383.
 SEQ ID NO:542 is the determined cDNA sequence for clone 25385.
 SEQ ID NO:543 is the determined cDNA sequence for clone 25386.
 SEQ ID NO:544 is the determined cDNA sequence for clone 25387.
- 5 SEQ ID NO:545 is the determined cDNA sequence for clone 26013.
 SEQ ID NO:546 is the determined cDNA sequence for clone 26014.
 SEQ ID NO:547 is the determined cDNA sequence for clone 26016.
 SEQ ID NO:548 is the determined cDNA sequence for clone 26017.
 SEQ ID NO:549 is the determined cDNA sequence for clone 26018.
- 10 SEQ ID NO:550 is the determined cDNA sequence for clone 26019.
 SEQ ID NO:551 is the determined cDNA sequence for clone 26020.
 SEQ ID NO:552 is the determined cDNA sequence for clone 26021.
 SEQ ID NO:553 is the determined cDNA sequence for clone 26022.
 SEQ ID NO:554 is the determined cDNA sequence for clone 26027.
- 15 SEQ ID NO:555 is the determined cDNA sequence for clone 26197.
 SEQ ID NO:556 is the determined cDNA sequence for clone 26199.
 SEQ ID NO:557 is the determined cDNA sequence for clone 26201.
 SEQ ID NO:558 is the determined cDNA sequence for clone 26202.
 SEQ ID NO:559 is the determined cDNA sequence for clone 26203.
- 20 SEQ ID NO:560 is the determined cDNA sequence for clone 26204.
 SEQ ID NO:561 is the determined cDNA sequence for clone 26205.
 SEQ ID NO:562 is the determined cDNA sequence for clone 26206.
 SEQ ID NO:563 is the determined cDNA sequence for clone 26208.
 SEQ ID NO:564 is the determined cDNA sequence for clone 26211.
- 25 SEQ ID NO:565 is the determined cDNA sequence for clone 26212.
 SEQ ID NO:566 is the determined cDNA sequence for clone 26213.
 SEQ ID NO:567 is the determined cDNA sequence for clone 26214.
 SEQ ID NO:568 is the determined cDNA sequence for clone 26215.
 SEQ ID NO:569 is the determined cDNA sequence for clone 26216.

- SEQ ID NO:570 is the determined cDNA sequence for clone 26217.
 SEQ ID NO:571 is the determined cDNA sequence for clone 26218.
 SEQ ID NO:572 is the determined cDNA sequence for clone 26219.
 SEQ ID NO:573 is the determined cDNA sequence for clone 26220.
 5 SEQ ID NO:574 is the determined cDNA sequence for clone 26221.
 SEQ ID NO:575 is the determined cDNA sequence for clone 26224.
 SEQ ID NO:576 is the determined cDNA sequence for clone 26225.
 SEQ ID NO:577 is the determined cDNA sequence for clone 26226.
 SEQ ID NO:578 is the determined cDNA sequence for clone 26227.
 10 SEQ ID NO:579 is the determined cDNA sequence for clone 26228.
 SEQ ID NO:580 is the determined cDNA sequence for clone 26230.
 SEQ ID NO:581 is the determined cDNA sequence for clone 26231.
 SEQ ID NO:582 is the determined cDNA sequence for clone 26234.
 SEQ ID NO:583 is the determined cDNA sequence for clone 26236.
 15 SEQ ID NO:584 is the determined cDNA sequence for clone 26237.
 SEQ ID NO:585 is the determined cDNA sequence for clone 26239.
 SEQ ID NO:586 is the determined cDNA sequence for clone 26240.
 SEQ ID NO:587 is the determined cDNA sequence for clone 26241.
 SEQ ID NO:588 is the determined cDNA sequence for clone 26242.
 20 SEQ ID NO:589 is the determined cDNA sequence for clone 26246.
 SEQ ID NO:590 is the determined cDNA sequence for clone 26247.
 SEQ ID NO:591 is the determined cDNA sequence for clone 26248.
 SEQ ID NO:592 is the determined cDNA sequence for clone 26249.
 SEQ ID NO:593 is the determined cDNA sequence for clone 26250.
 25 SEQ ID NO:594 is the determined cDNA sequence for clone 26251.
 SEQ ID NO:595 is the determined cDNA sequence for clone 26252.
 SEQ ID NO:596 is the determined cDNA sequence for clone 26253.
 SEQ ID NO:597 is the determined cDNA sequence for clone 26254.
 SEQ ID NO:598 is the determined cDNA sequence for clone 26255.

- SEQ ID NO:599 is the determined cDNA sequence for clone 26256.
 SEQ ID NO:600 is the determined cDNA sequence for clone 26257.
 SEQ ID NO:601 is the determined cDNA sequence for clone 26259.
 SEQ ID NO:602 is the determined cDNA sequence for clone 26260.
 5 SEQ ID NO:603 is the determined cDNA sequence for clone 26261.
 SEQ ID NO:604 is the determined cDNA sequence for clone 26262.
 SEQ ID NO:605 is the determined cDNA sequence for clone 26263.
 SEQ ID NO:606 is the determined cDNA sequence for clone 26264.
 SEQ ID NO:607 is the determined cDNA sequence for clone 26265.
 10 SEQ ID NO:608 is the determined cDNA sequence for clone 26266.
 SEQ ID NO:609 is the determined cDNA sequence for clone 26268.
 SEQ ID NO:610 is the determined cDNA sequence for clone 26269.
 SEQ ID NO:611 is the determined cDNA sequence for clone 26271.
 SEQ ID NO:612 is the determined cDNA sequence for clone 26273.
 15 SEQ ID NO:613 is the determined cDNA sequence for clone 26810.
 SEQ ID NO:614 is the determined cDNA sequence for clone 26811.
 SEQ ID NO:615 is the determined cDNA sequence for clone 26812.1.
 SEQ ID NO:616 is the determined cDNA sequence for clone 26812.2.
 SEQ ID NO:617 is the determined cDNA sequence for clone 26813.
 20 SEQ ID NO:618 is the determined cDNA sequence for clone 26814.
 SEQ ID NO:619 is the determined cDNA sequence for clone 26815.
 SEQ ID NO:620 is the determined cDNA sequence for clone 26816.
 SEQ ID NO:621 is the determined cDNA sequence for clone 26818.
 SEQ ID NO:622 is the determined cDNA sequence for clone 26819.
 25 SEQ ID NO:623 is the determined cDNA sequence for clone 26820.
 SEQ ID NO:624 is the determined cDNA sequence for clone 26821.
 SEQ ID NO:625 is the determined cDNA sequence for clone 26822.
 SEQ ID NO:626 is the determined cDNA sequence for clone 26824.
 SEQ ID NO:627 is the determined cDNA sequence for clone 26825.

- SEQ ID NO:628 is the determined cDNA sequence for clone 26826.
 SEQ ID NO:629 is the determined cDNA sequence for clone 26827.
 SEQ ID NO:630 is the determined cDNA sequence for clone 26829.
 SEQ ID NO:631 is the determined cDNA sequence for clone 26830.
 5 SEQ ID NO:632 is the determined cDNA sequence for clone 26831.
 SEQ ID NO:633 is the determined cDNA sequence for clone 26832.
 SEQ ID NO:634 is the determined cDNA sequence for clone 26835.
 SEQ ID NO:635 is the determined cDNA sequence for clone 26836.
 SEQ ID NO:636 is the determined cDNA sequence for clone 26837.
 10 SEQ ID NO:637 is the determined cDNA sequence for clone 26839.
 SEQ ID NO:638 is the determined cDNA sequence for clone 26841.
 SEQ ID NO:639 is the determined cDNA sequence for clone 26843.
 SEQ ID NO:640 is the determined cDNA sequence for clone 26844.
 SEQ ID NO:641 is the determined cDNA sequence for clone 26845.
 15 SEQ ID NO:642 is the determined cDNA sequence for clone 26846.
 SEQ ID NO:643 is the determined cDNA sequence for clone 26847.
 SEQ ID NO:644 is the determined cDNA sequence for clone 26848.
 SEQ ID NO:645 is the determined cDNA sequence for clone 26849.
 SEQ ID NO:646 is the determined cDNA sequence for clone 26850.
 20 SEQ ID NO:647 is the determined cDNA sequence for clone 26851.
 SEQ ID NO:648 is the determined cDNA sequence for clone 26852.
 SEQ ID NO:649 is the determined cDNA sequence for clone 26853.
 SEQ ID NO:650 is the determined cDNA sequence for clone 26854.
 SEQ ID NO:651 is the determined cDNA sequence for clone 26856.
 25 SEQ ID NO:652 is the determined cDNA sequence for clone 26857.
 SEQ ID NO:653 is the determined cDNA sequence for clone 26858.
 SEQ ID NO:654 is the determined cDNA sequence for clone 26859.
 SEQ ID NO:655 is the determined cDNA sequence for clone 26860.
 SEQ ID NO:656 is the determined cDNA sequence for clone 26862.

- SEQ ID NO:657 is the determined cDNA sequence for clone 26863.
 SEQ ID NO:658 is the determined cDNA sequence for clone 26864.
 SEQ ID NO:659 is the determined cDNA sequence for clone 26865.
 SEQ ID NO:660 is the determined cDNA sequence for clone 26867.
- 5 SEQ ID NO:661 is the determined cDNA sequence for clone 26868.
 SEQ ID NO:662 is the determined cDNA sequence for clone 26871.
 SEQ ID NO:663 is the determined cDNA sequence for clone 26873.
 SEQ ID NO:664 is the determined cDNA sequence for clone 26875.
 SEQ ID NO:665 is the determined cDNA sequence for clone 26876.
- 10 SEQ ID NO:666 is the determined cDNA sequence for clone 26877.
 SEQ ID NO:667 is the determined cDNA sequence for clone 26878.
 SEQ ID NO:668 is the determined cDNA sequence for clone 26880.
 SEQ ID NO:669 is the determined cDNA sequence for clone 26882.
 SEQ ID NO:670 is the determined cDNA sequence for clone 26883.
- 15 SEQ ID NO:671 is the determined cDNA sequence for clone 26884.
 SEQ ID NO:672 is the determined cDNA sequence for clone 26885.
 SEQ ID NO:673 is the determined cDNA sequence for clone 26886.
 SEQ ID NO:674 is the determined cDNA sequence for clone 26887.
 SEQ ID NO:675 is the determined cDNA sequence for clone 26888.
- 20 SEQ ID NO:676 is the determined cDNA sequence for clone 26889.
 SEQ ID NO:677 is the determined cDNA sequence for clone 26890.
 SEQ ID NO:678 is the determined cDNA sequence for clone 26892.
 SEQ ID NO:679 is the determined cDNA sequence for clone 26894.
 SEQ ID NO:680 is the determined cDNA sequence for clone 26895.
- 25 SEQ ID NO:681 is the determined cDNA sequence for clone 26897.
 SEQ ID NO:682 is the determined cDNA sequence for clone 26898.
 SEQ ID NO:683 is the determined cDNA sequence for clone 26899.
 SEQ ID NO:684 is the determined cDNA sequence for clone 26900.
 SEQ ID NO:685 is the determined cDNA sequence for clone 26901.

- SEQ ID NO:686 is the determined cDNA sequence for clone 26903.
 SEQ ID NO:687 is the determined cDNA sequence for clone 26905.
 SEQ ID NO:688 is the determined cDNA sequence for clone 26906.
 SEQ ID NO:689 is the determined cDNA sequence for clone 26708.
 5 SEQ ID NO:690 is the determined cDNA sequence for clone 26709.
 SEQ ID NO:691 is the determined cDNA sequence for clone 26710.
 SEQ ID NO:692 is the determined cDNA sequence for clone 26711.
 SEQ ID NO:693 is the determined cDNA sequence for clone 26712.
 SEQ ID NO:694 is the determined cDNA sequence for clone 26713.
 10 SEQ ID NO:695 is the determined cDNA sequence for clone 26714.
 SEQ ID NO:696 is the determined cDNA sequence for clone 26715.
 SEQ ID NO:697 is the determined cDNA sequence for clone 26716.
 SEQ ID NO:698 is the determined cDNA sequence for clone 26717.
 SEQ ID NO:699 is the determined cDNA sequence for clone 26718.
 15 SEQ ID NO:700 is the determined cDNA sequence for clone 26719.
 SEQ ID NO:701 is the determined cDNA sequence for clone 26720.
 SEQ ID NO:702 is the determined cDNA sequence for clone 26721.
 SEQ ID NO:703 is the determined cDNA sequence for clone 26722.
 SEQ ID NO:704 is the determined cDNA sequence for clone 26723.
 20 SEQ ID NO:705 is the determined cDNA sequence for clone 26724.
 SEQ ID NO:706 is the determined cDNA sequence for clone 26725.
 SEQ ID NO:707 is the determined cDNA sequence for clone 26726.
 SEQ ID NO:708 is the determined cDNA sequence for clone 26727.
 SEQ ID NO:709 is the determined cDNA sequence for clone 26728.
 25 SEQ ID NO:710 is the determined cDNA sequence for clone 26729.
 SEQ ID NO:711 is the determined cDNA sequence for clone 26730.
 SEQ ID NO:712 is the determined cDNA sequence for clone 26731.
 SEQ ID NO:713 is the determined cDNA sequence for clone 26732.
 SEQ ID NO:714 is the determined cDNA sequence for clone 26733.1.

- SEQ ID NO:715 is the determined cDNA sequence for clone 26733.2.
- SEQ ID NO:716 is the determined cDNA sequence for clone 26734.
- SEQ ID NO:717 is the determined cDNA sequence for clone 26735.
- SEQ ID NO:718 is the determined cDNA sequence for clone 26736.
- 5 SEQ ID NO:719 is the determined cDNA sequence for clone 26737.
- SEQ ID NO:720 is the determined cDNA sequence for clone 26738.
- SEQ ID NO:721 is the determined cDNA sequence for clone 26739.
- SEQ ID NO:722 is the determined cDNA sequence for clone 26741.
- SEQ ID NO:723 is the determined cDNA sequence for clone 26742.
- 10 SEQ ID NO:724 is the determined cDNA sequence for clone 26743.
- SEQ ID NO:725 is the determined cDNA sequence for clone 26744.
- SEQ ID NO:726 is the determined cDNA sequence for clone 26745.
- SEQ ID NO:727 is the determined cDNA sequence for clone 26746.
- SEQ ID NO:728 is the determined cDNA sequence for clone 26747.
- 15 SEQ ID NO:729 is the determined cDNA sequence for clone 26748.
- SEQ ID NO:730 is the determined cDNA sequence for clone 26749.
- SEQ ID NO:731 is the determined cDNA sequence for clone 26750.
- SEQ ID NO:732 is the determined cDNA sequence for clone 26751.
- SEQ ID NO:733 is the determined cDNA sequence for clone 26752.
- 20 SEQ ID NO:734 is the determined cDNA sequence for clone 26753.
- SEQ ID NO:735 is the determined cDNA sequence for clone 26754.
- SEQ ID NO:736 is the determined cDNA sequence for clone 26755.
- SEQ ID NO:737 is the determined cDNA sequence for clone 26756.
- SEQ ID NO:738 is the determined cDNA sequence for clone 26757.
- 25 SEQ ID NO:739 is the determined cDNA sequence for clone 26758.
- SEQ ID NO:740 is the determined cDNA sequence for clone 26759.
- SEQ ID NO:741 is the determined cDNA sequence for clone 26760.
- SEQ ID NO:742 is the determined cDNA sequence for clone 26761.
- SEQ ID NO:743 is the determined cDNA sequence for clone 26762.

- SEQ ID NO:744 is the determined cDNA sequence for clone 26763.
 SEQ ID NO:745 is the determined cDNA sequence for clone 26764.
 SEQ ID NO:746 is the determined cDNA sequence for clone 26765.
 SEQ ID NO:747 is the determined cDNA sequence for clone 26766.
 5 SEQ ID NO:748 is the determined cDNA sequence for clone 26767.
 SEQ ID NO:749 is the determined cDNA sequence for clone 26768.
 SEQ ID NO:750 is the determined cDNA sequence for clone 26769.
 SEQ ID NO:751 is the determined cDNA sequence for clone 26770.
 SEQ ID NO:752 is the determined cDNA sequence for clone 26771.
 10 SEQ ID NO:753 is the determined cDNA sequence for clone 26772.
 SEQ ID NO:754 is the determined cDNA sequence for clone 26773.
 SEQ ID NO:755 is the determined cDNA sequence for clone 26774.
 SEQ ID NO:756 is the determined cDNA sequence for clone 26775.
 SEQ ID NO:757 is the determined cDNA sequence for clone 26776.
 15 SEQ ID NO:758 is the determined cDNA sequence for clone 26777.
 SEQ ID NO:759 is the determined cDNA sequence for clone 26778.
 SEQ ID NO:760 is the determined cDNA sequence for clone 26779.
 SEQ ID NO:761 is the determined cDNA sequence for clone 26781.
 SEQ ID NO:762 is the determined cDNA sequence for clone 26782.
 20 SEQ ID NO:763 is the determined cDNA sequence for clone 26783.
 SEQ ID NO:764 is the determined cDNA sequence for clone 26784.
 SEQ ID NO:765 is the determined cDNA sequence for clone 26785.
 SEQ ID NO:766 is the determined cDNA sequence for clone 26786.
 SEQ ID NO:767 is the determined cDNA sequence for clone 26787.
 25 SEQ ID NO:768 is the determined cDNA sequence for clone 26788.
 SEQ ID NO:769 is the determined cDNA sequence for clone 26790.
 SEQ ID NO:770 is the determined cDNA sequence for clone 26791.
 SEQ ID NO:771 is the determined cDNA sequence for clone 26792.
 SEQ ID NO:772 is the determined cDNA sequence for clone 26793.

- SEQ ID NO:773 is the determined cDNA sequence for clone 26794.
 SEQ ID NO:774 is the determined cDNA sequence for clone 26795.
 SEQ ID NO:775 is the determined cDNA sequence for clone 26796.
 SEQ ID NO:776 is the determined cDNA sequence for clone 26797.
 5 SEQ ID NO:777 is the determined cDNA sequence for clone 26798.
 SEQ ID NO:778 is the determined cDNA sequence for clone 26800.
 SEQ ID NO:779 is the determined cDNA sequence for clone 26801.
 SEQ ID NO:780 is the determined cDNA sequence for clone 26802.
 SEQ ID NO:781 is the determined cDNA sequence for clone 26803.
 10 SEQ ID NO:782 is the determined cDNA sequence for clone 26804.
 SEQ ID NO:783 is the amino acid sequence for L773P.
 SEQ ID NO:784 is the determined DNA sequence of the L773P expression construct.
 SEQ ID NO:785 is the determined DNA sequence of the L773PA expression construct.
 SEQ ID NO:786 is a predicted amino acid sequence for L552S.
 15 SEQ ID NO:787 is a predicted amino acid sequence for L840P.
 SEQ ID NO:788 is the full-length cDNA sequence for L548S.
 SEQ ID NO:789 is the amino acid sequence encoded by SEQ ID NO:788.
 SEQ ID NO:790 is an extended cDNA sequence for L552S.
 SEQ ID NO:791 is the predicted amino acid sequence encoded by the cDNA sequence of
 20 SEQ ID NO:790.
 SEQ ID NO:792 is the determined cDNA sequence for an isoform of L552S.
 SEQ ID NO:793 is the predicted amino acid sequence encoded by SEQ ID NO:792.
 SEQ ID NO:794 is an extended cDNA sequence for L840P.
 SEQ ID NO:795 is the predicted amino acid sequence encoded by SEQ DI NO:794.
 25 SEQ ID NO:796 is an extended cDNA sequence for L801P.
 SEQ ID NO:797 is a first predicted amino acid sequence encoded by SEQ ID NO:796.
 SEQ ID NO:798 is a second predicted amino acid sequence encoded by SEQ ID NO:796.
 SEQ ID NO:799 is a third predicted amino acid sequence encoded by SEQ ID NO:796.
 SEQ ID NO:800 is the determined full-length sequence for L844P.

- SEQ ID NO:801 is the 5' consensus cDNA sequence for L551S.
 SEQ ID NO:802 is the 3' consensus cDNA sequence for L551S.
 SEQ ID NO:803 is the cDNA sequence for STY8.
 SEQ ID NO:804 is an extended cDNA sequence for L551S.
- 5 SEQ ID NO:805 is the amino acid sequence for STY8.
 SEQ ID NO:806 is the extended amino acid sequence for L551S.
 SEQ ID NO:807 is the determined full length cDNA sequence for L773P.
 SEQ ID NO:808 is the full-length cDNA sequence of L552S.
 SEQ ID NO:809 is the full-length amino acid sequence of L552S.
- 10 SEQ ID NO:810 is the determined cDNA sequence of clone 50989.
 SEQ ID NO:811 is the determined cDNA sequence of clone 50990.
 SEQ ID NO:812 is the determined cDNA sequence of clone 50992.
 SEQ ID NO:813-824 are the determined cDNA sequences for clones isolated from lung tumor tissue.
- 15 SEQ ID NO:825 is the determined cDNA sequence for the full-length L551S clone 54305.
 SEQ ID NO:826 is the determined cDNA sequence for the full-length L551S clone 54298.
 SEQ ID NO:827 is the full-length amino acid sequence for L551S.

Tables 1-6 contain the sequence identifiers for SEQ ID NO:828-1664.

Table 1A

5

SEQ ID NO:	CLONE IDENTIFIER	SEQ ID NO:	CLONE IDENTIFIER
828	R0126:A02	869	R0126:D12
829	R0126:A03	870	R0126:E01
830	R0126:A05	871	R0126:E02
831	R0126:A06	872	R0126:E03
832	R0126:A08	873	R0126:E04
833	R0126:A09	874	R0126:E05
834	R0126:A10	875	R0126:E06
835	R0126:A11	876	R0126:E07
836	R0126:A12	877	R0126:E08
837	R0126:B01	878	R0126:E09
838	R0126:B03	879	R0126:E10
839	R0126:B04	880	R0126:E11
840	R0126:B05	881	R0126:E12
841	R0126:B06	882	R0126:F01
842	R0126:B07	883	R0126:F02
843	R0126:B08	884	R0126:F03
844	R0126:B09	885	R0126:F04
845	R0126:B11	886	R0126:F05
846	R0126:B12	887	R0126:F06
847	R0126:C01	888	R0126:F07
848	R0126:C02	889	R0126:F08
849	R0126:C03	890	R0126:F10
850	R0126:C05	891	R0126:F11
851	R0126:C06	892	R0126:F12
852	R0126:C07	893	R0126:G01
853	R0126:C08	894	R0126:G02
854	R0126:C09	895	R0126:G03
855	R0126:C10	896	R0126:G04
856	R0126:C11	897	R0126:G05
857	R0126:C12	898	R0126:G06
858	R0126:D01	899	R0126:G07
859	R0126:D02	900	R0126:G09
860	R0126:D03	901	R0126:G10
861	R0126:D04	902	R0126:G11

SEQ ID NO:	CLONE IDENTIFIER	SEQ ID NO:	CLONE IDENTIFIER
862	R0126:D05	903	R0126:G12
863	R0126:D06	904	R0126:H01
864	R0126:D07	905	R0126:H02
865	R0126:D08	906	R0126:H03
866	R0126:D09	907	R0126:H04
867	R0126:D10	908	R0126:H05
868	R0126:D11	909	R0126:H06

bioRxiv preprint doi: <https://doi.org/10.1101/147817>; this version posted November 1, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1B

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
910	R0126:H07	951	R0127:D10
911	R0126:H09	952	R0127:D11
912	R0126:H10	953	R0127:D12
913	R0126:H11	954	R0127:E02
914	R0127:A02	955	R0127:E03
915	R0127:A05	956	R0127:E04
916	R0127:A06	957	R0127:E05
917	R0127:A07	958	R0127:E06
918	R0127:A08	959	R0127:E07
919	R0127:A09	960	R0127:E08
920	R0127:A10	961	R0127:E09
921	R0127:A11	962	R0127:E10
922	R0127:A12	963	R0127:E11
923	R0127:B01	964	R0127:F01
924	R0127:B03	965	R0127:F02
925	R0127:B04	966	R0127:F03
926	R0127:B05	967	R0127:F04
927	R0127:B06	968	R0127:F05
928	R0127:B07	969	R0127:F06
929	R0127:B08	970	R0127:F07
930	R0127:B09	971	R0127:F08
931	R0127:B10	972	R0127:F10
932	R0127:B11	973	R0127:F11
933	R0127:B12	974	R0127:F12
934	R0127:C01	975	R0127:G01
935	R0127:C03	976	R0127:G02
936	R0127:C04	977	R0127:G03
937	R0127:C05	978	R0127:G04
938	R0127:C07	979	R0127:G05
939	R0127:C08	980	R0127:G06
940	R0127:C09	981	R0127:G07
941	R0127:C10	982	R0127:G08
942	R0127:C11	983	R0127:G09
943	R0127:D01	984	R0127:G10
944	R0127:D02	985	R0127:G11
945	R0127:D03	986	R0127:G12
946	R0127:D04	987	R0127:H01
947	R0127:D05	988	R0127:H02
948	R0127:D06	989	R0127:H03

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
949	R0127:D07	990	R0127:H04
950	R0127:D01	991	R0127:H05

BIOINFORMATICS

Table 1C

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
992	R0127:H06	1034	R0128:D11
993	R0127:H07	1035	R0128:D12
994	R0127:H08	1036	R0128:E01
995	R1027:H09	1037	R0128:E02
996	R1027:H10	1038	R0128:E03
997	R1027:H11	1039	R0128:E04
998	R1028:A02	1040	R0128:E05
999	R1028:A05	1041	R0128:E06
1000	R1028:A06	1042	R0128:E07
1001	R1028:A07	1043	R0128:E08
1002	R1028:A08	1044	R0128:E09
1003	R1028:A09	1045	R0128:E10
1004	R1028:A10	1046	R0128:E12
1005	R1028:B01	1047	R0128:F01
1006	R1028:B02	1048	R0128:F02
1007	R1028:B03	1049	R0128:F03
1008	R1028:B04	1050	R0128:F04
1009	R1028:B05	1051	R0128:F06
1010	R1028:B08	1052	R0128:F07
1011	R1028:B09	1053	R0128:F08
1012	R1028:B10	1054	R0128:F09
1013	R1028:B11	1055	R0128:F10
1014	R1028:B12	1056	R0128:F12
1015	R1028:C01	1057	R0128:G01
1016	R1028:C03	1058	R0128:G02
1017	R1028:C04	1059	R0128:G03
1018	R1028:C05	1060	R0128:G04
1019	R1028:C06	1061	R0128:G05
1020	R1028:C07	1062	R0128:G06
1021	R1028:C08	1063	R0128:G07
1022	R1028:C10	1064	R0128:G09
1023	R1028:C11	1065	R0128:G10
1024	R1028:C12	1066	R0128:G11
1025	R1028:D01	1067	R0128:G12
1026	R1028:D02	1068	R0128:H01
1027	R1028:D04	1069	R0128:H02
1028	R1028:D05	1070	R0128:H03
1029	R1028:D06	1071	R0128:H04
1030	R1028:D07	1072	R0128:H05

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1031	R1028:D08	1073	R0128:H06
1032	R1028:D09	1074	R0128:H07
1033	R0128:D10	1075	R0128:H08

R0128:H06

Table 1D

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1076	R0128:H09	1118	R0130:D12
1077	R0128:H10	1119	R0130:E01
1078	R0128:H11	1120	R0130:E02
1079	R0130:A02	1121	R0130:E03
1080	R0130:A05	1122	R0130:E04
1081	R0130:A06	1123	R0130:E05
1082	R0130:A08	1124	R0130:E06
1083	R0130:A09	1125	R0130:E07
1084	R0130:A10	1126	R0130:E08
1085	R0130:A11	1127	R0130:E09
1086	R0130:A12	1128	R0130:E10
1087	R0130:B01	1129	R0130:E11
1088	R0130:B02	1130	R0130:E12
1089	R0130:B03	1131	R0130:F02
1090	R0130:B04	1132	R0130:F03
1091	R0130:B05	1133	R0130:F05
1092	R0130:B06	1134	R0130:F06
1093	R0130:B08	1135	R0130:F07
1094	R0130:B09	1136	R0130:F08
1095	R0130:B10	1137	R0130:F09
1096	R0130:B11	1138	R0130:F10
1097	R0130:B12	1139	R0130:F11
1098	R0130:C02	1140	R0130:F12
1099	R0130:C03	1141	R0130:G01
1100	R0130:C04	1142	R0130:G02
1101	R0130:C05	1143	R0130:G03
1102	R0130:C06	1144	R0130:G04
1103	R0130:C07	1145	R0130:G05
1104	R0130:C08	1146	R0130:G06
1105	R0130:C09	1147	R0130:G07
1106	R0130:C10	1148	R0130:G08
1107	R0130:C11	1149	R0130:G09
1108	R0130:C12	1150	R0130:G10
1109	R0130:D02	1151	R0130:G11
1110	R0130:D03	1152	R0130:G12
1111	R0130:D04	1153	R0130:H01
1112	R0130:D05	1154	R0130:H02
1113	R0130:D06	1155	R0130:H04
1114	R0130:D07	1156	R0130:H05

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1115	R0130:D09	1157	R0130:H06
1116	R0130:D10	1158	R0130:H07
1117	R0130:D11	1159	R0130:H08

Table 1E

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1160	R0130:H09	1200	R0131:E01
1161	R0130:H10	1201	R0131:E02
1162	R0130:H11	1202	R0131:E03
1163	R0131:A02	1203	R0131:E04
1164	R0131:A05	1204	R0131:E06
1165	R0131:A06	1205	R0131:E07
1166	R0131:A07	1206	R0131:E08
1167	R0131:A08	1207	R0131:E10
1168	R0131:A09	1208	R0131:E11
1169	R0131:A11	1209	R0131:E12
1170	R0131:A12	1210	R0131:F02
1171	R0131:B02	1211	R0131:F04
1172	R0131:B03	1212	R0131:F05
1173	R0131:B04	1213	R0131:F06
1174	R0131:B05	1214	R0131:F07
1175	R0131:B07	1215	R0131:F08
1176	R0131:B08	1216	R0131:F09
1177	R0131:B09	1217	R0131:F10
1178	R0131:B10	1218	R0131:F11
1179	R0131:B11	1219	R0131:F12
1180	R0131:C01	1220	R0131:G01
1181	R0131:C02	1221	R0131:G02
1182	R0131:C03	1222	R0131:G03
1183	R0131:C04	1223	R0131:G04
1184	R0131:C06	1224	R0131:G05
1185	R0131:C07	1225	R0131:G06
1186	R0131:C08	1226	R0131:G07
1187	R0131:C10	1227	R0131:G08
1188	R0131:C11	1228	R0131:G09
1189	R0131:C12	1229	R0131:G10
1190	R0131:D02	1230	R0131:G11
1191	R0131:D03	1231	R0131:G12
1192	R0131:D04	1232	R0131:H01
1193	R0131:D05	1233	R0131:H02
1194	R0131:D06	1234	R0131:H05
1195	R0131:D07	1235	R0131:H06
1196	R0131:D09	1236	R0131:H07
1197	R0131:D10	1237	R0131:H08
1198	R0131:D11	1238	R0131:H09

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1199	R0131:D12	1239	R0131:H11

R0131:H11

Table 2:
Clone names for NSCLC-SQL1 and corresponding SEQ ID NOs

SEQ ID NO	CLONE IDENTIFIER
1240	Contig 54
1241	Contig 55
1242	Contig 57
1243	Contig 58
1244	Contig 60
1245	Contig 62
1246	Contig 63
1247	Contig 64
1248	Contig 65
1249	Contig 66
1250	Contig 67
1251	Contig 68
1252	Contig 69
1253	Contig 70
1254	Contig 71
1255	Contig 72
1256	Contig 73
1257	Contig 74
1258	Contig 75
1259	Contig 77
1260	Contig 78
1261	Contig 79
1262	Contig 80
1263	Contig 81
1264	Contig 83
1265	Contig 84
1266	Contig 86
1267	Contig 87
1268	Contig 88
1269	Contig 89
1270	Contig 90
1271	Contig 91
1272	Contig 92
1273	Contig 94
1274	Contig 95
1275	Contig 96
1276	Contig 97
1277	Contig 98

SEQ ID NO	CLONE IDENTIFIER
1278	Contig 99
1279	Contig 100

210121.478C17

Table 3:
Clone names for NSCLC-SCLI and corresponding SEQ ID NOs

SEQ ID NO	CLONE IDENTIFIER
1280	Contig 38
1281	Contig 39
1282	Contig 40
1283	Contig 41
1284	Contig 42
1285	Contig 43
1286	Contig 44
1287	Contig 45
1288	Contig 46
1289	Contig 47
1290	Contig 48
1291	Contig 49
1292	Contig 51
1293	Contig 52
1294	Contig 53
1295	Contig 54
1296	Contig 55
1297	Contig 56
1298	Contig 57
1299	Contig 58
1300	Contig 59
1301	Contig 60
1302	Contig 62
1303	Contig 63
1304	Contig 64
1305	Contig 65
1306	Contig 66
1307	Contig 67
1308	Contig 68
1309	Contig 69
1310	Contig 70
1311	Contig 72
1312	Contig 73
1313	Contig 75
1314	Contig 76
1315	Contig 77
1316	Contig 78
1317	Contig 79

SEQ ID NO	CLONE IDENTIFIER
1318	Contig 80
1319	Contig 81
1320	Contig 82

Contig 80

Table 4A:
Clone names for NSCLC-SCL3-SCL4 and corresponding SEQ ID NOs

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1321	Contig 94	1363	Contig 136
1322	Contig 95	1364	Contig 137
1323	Contig 96	1365	Contig 138
1324	Contig 97	1366	Contig 139
1325	Contig 98	1367	Contig 140
1326	Contig 99	1368	Contig 141
1327	Contig 100	1369	Contig 142
1328	Contig 101	1370	Contig 143
1329	Contig 102	1371	Contig 144
1330	Contig 103	1372	Contig 145
1331	Contig 104	1373	Contig 146
1332	Contig 105	1374	Contig 147
1333	Contig 106	1375	Contig 148
1334	Contig 107	1376	Contig 149
1335	Contig 108	1377	Contig 150
1336	Contig 109	1378	Contig 151
1337	Contig 110	1379	Contig 152
1338	Contig 111	1380	Contig 153
1339	Contig 112	1381	Contig 154
1340	Contig 113	1382	Contig 155
1341	Contig 114	1383	Contig 156
1342	Contig 115	1384	Contig 157
1343	Contig 116	1385	Contig 158
1344	Contig 117	1386	Contig 159
1345	Contig 118	1387	Contig 160
1346	Contig 119	1388	Contig 161
1347	Contig 120	1389	Contig 162
1348	Contig 121	1390	Contig 163
1349	Contig 122	1391	Contig 164
1350	Contig 123	1392	Contig 165
1351	Contig 124	1393	Contig 166
1352	Contig 125	1394	Contig 167
1353	Contig 126	1395	Contig 168
1354	Contig 127	1396	Contig 169
1355	Contig 128	1397	Contig 170
1356	Contig 129	1398	Contig 171
1357	Contig 130	1399	Contig 172
1358	Contig 131	1400	Contig 173

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1359	Contig 132	1401	Contig 174
1360	Contig 133	1402	Contig 175
1361	Contig 134	1403	Contig 176
1362	Contig 135		

Table 4B:

5 **Clone names for NSCLC-SCL3-SCL4 and corresponding SEQ ID NOs**

SEQ ID NO	CLONE IDENTIFIER
1404	Contig 177
1405	Contig 178
1406	Contig 179
1407	Contig 180
1408	Contig 181
1409	Contig 182
1410	Contig 183
1411	Contig 184
1412	Contig 185
1413	Contig 186
1414	Contig 187

Table 5:
Clone names for SCLC-SQL1 and corresponding SEQ ID NOs

SEQ ID NO	CLONE IDENTIFIER
1415	Contig 17
1416	Contig 18
1417	Contig 20
1418	Contig 23
1419	Contig 24
1420	Contig 25
1421	Contig 26
1422	Contig 27
1423	Contig 28
1424	Contig 29
1425	Contig 30
1426	Contig 31
1427	Contig 20
1428	Contig 21
1429	Contig 22
1430	Contig 23
1431	Contig 24
1432	Contig 25
1433	Contig 26
1434	Contig 27
1435	Contig 28
1436	Contig 29
1437	Contig 30
1438	Contig 31
1439	Contig 32
1440	Contig 33
1441	Contig 34
1442	Contig 35
1443	Contig 36
1444	Contig 37
1445	Contig 38

Table 6A:
Clone names for SCLC-SCL3-SCL4 and corresponding SEQ ID NOs

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1446	Contig 116	1488	Contig 160
1447	Contig 117	1489	Contig 161
1448	Contig 118	1490	Contig 162
1449	Contig 119	1491	Contig 163
1450	Contig 120	1492	Contig 164
1451	Contig 122	1493	Contig 165
1452	Contig 123	1494	Contig 166
1453	Contig 124	1495	Contig 167
1454	Contig 125	1496	Contig 168
1455	Contig 126	1497	Contig 169
1456	Contig 127	1498	Contig 170
1457	Contig 128	1499	Contig 171
1458	Contig 129	1500	Contig 172
1459	Contig 130	1501	Contig 173
1460	Contig 131	1502	Contig 174
1461	Contig 132	1503	Contig 175
1462	Contig 133	1504	Contig 176
1463	Contig 135	1505	Contig 177
1464	Contig 136	1506	Contig 178
1465	Contig 137	1507	Contig 179
1466	Contig 138	1508	Contig 181
1467	Contig 139 (L985P)	1509	Contig 182
1468	Contig 140	1510	Contig 183
1469	Contig 141	1511	Contig 184
1470	Contig 142	1512	Contig 185
1471	Contig 143	1513	Contig 186
1472	Contig 144	1514	Contig 187
1473	Contig 145	1515	Contig 189
1474	Contig 146	1516	Contig 190
1475	Contig 147	1517	Contig 191
1476	Contig 148	1518	Contig 192
1477	Contig 149	1519	Contig 193
1478	Contig 150	1520	Contig 194
1479	Contig 151	1521	Contig 195
1480	Contig 152	1522	Contig 196
1481	Contig 153	1523	Contig 197
1482	Contig 154	1524	Contig 198
1483	Contig 155	1525	Contig 199

SEQ ID NO	CLONE IDENTIFIER	SEQ ID NO	CLONE IDENTIFIER
1484	Contig 156	1526	Contig 200
1485	Contig 157	1527	Contig 201
1486	Contig 158	1528	Contig 202
1487	Contig 159		

bioRxiv preprint doi: <https://doi.org/10.1101/148484>; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 6B:
Clone names for SCLC-SCL3-SCL4 and corresponding SEQ ID NOs

SEQ ID NO	CLONE IDENTIFIER
1529	Contig 203
1530	Contig 204
1531	Contig 205
1532	Contig 206
1533	Contig 207
1534	Contig 208
1535	Contig 209
1536	Contig 210
1537	Contig 211
1538	Contig 212
1539	Contig 213
1540	Contig 214
1541	Contig 215
1542	Contig 216
1543	Contig 217
1544	Contig 218
1545	Contig 219
1546	Contig 220
1547	Contig 221
1548	Contig 222
1549	Contig 223
1550	Contig 224
1551	Contig 225
1552	Contig 226
1553	Contig 227
1554	Contig 228
1555	Contig 229
1556	Contig 230
1557	Contig 231
1558	Contig 232
1559	Contig 233
1560	Contig 234
1561	Contig 235
1562	Contig 236
1563	Contig 237

Table 7.

SEQ ID NO:	CLONE IDENTIFIER	SEQ ID NO:	CLONE IDENTIFIER
1564	R0124:E05	1609	R0129:D09
1565	R0124:E06	1610	R0129:D10
1566	R0124:E08	1611	R0129:D11
1567	R0124:F07	1612	R0129:E02
1568	R0124:F08	1613	R0129:E03
1569	R0124:F09	1614	R0129:E04
1570	R0124:G04	1615	R0129:E05
1571	R0129:A02	1616	R0129:E06
1572	R0129:A03	1617	R0129:E07
1573	R0129:A06	1618	R0129:E08
1574	R0129:A07	1619	R0129:E09
1575	R0129:A08	1620	R0129:E11
1576	R0129:A09	1621	R0129:E12
1577	R0129:A10	1622	R0129:F01
1578	R0129:A11	1623	R0129:F02
1579	R0129:A12	1624	R0129:F03
1580	R0129:B02	1625	R0129:F04
1581	R0129:B03	1626	R0129:F06
1582	R0129:B04	1627	R0129:F07
1583	R0129:B05	1628	R0129:F08
1584	R0129:B06	1629	R0129:F09
1585	R0129:B07	1630	R0129:F10
1586	R0129:B08	1631	R0129:F11
1587	R0129:B09	1632	R0129:F12
1588	R0129:B10	1633	R0129:G01
1589	R0129:B11	1634	R0129:G02
1590	R0129:B12	1635	R0129:G03
1591	R0129:C01	1636	R0129:G04
1592	R0129:C02	1637	R0129:G05
1593	R0129:C03	1638	R0129:G06
1594	R0129:C04	1639	R0129:G07
1595	R0129:C06	1640	R0129:G08
1596	R0129:C07	1641	R0129:G09
1597	R0129:C08	1642	R0129:G10
1598	R0129:C09	1643	R0129:G11
1599	R0129:C10	1644	R0129:G12
1600	R0129:C11	1645	R0129:H01
1601	R0129:C12	1646	R0129:H02
1602	R0129:D01	1647	R0129:H03

1603	R0129:D03	1648	R0129:H04
1604	R0129:D04	1649	R0129:H05
1605	R0129:D05	1650	R0129:H08
1606	R0129:D06	1651	R0129:H09
1607	R0129:D07	1652	R0129:H10
1608	R0129:D08	1653	R0129:H11

Table 8

5

SEQ ID NO:	CLONE IDENTIFIER
1654	26484
1655	26496
1656	26517
1657	26531
1658	26022
1659	26026
1660	26810
1661	26815
1662	26869
1663	26883
1664	26902

SEQ ID NO:1665 and 1666 are primers used in the amplification of the coding region of L548S

SEQ ID NO:1667 is the protein sequence of expressed recombinant L7548S.

10 SEQ ID NO:1668 is the cDNA sequence of expressed recombinant L7548S.

SEQ ID NO:1669 is the extended cDNA sequence of clone #18971 (L801P).

SEQ ID NO:1670 is the amino acid sequence of open reading frame ORF4 encoded by SEQ ID NO:1669.

SEQ ID NO:1671 is the amino acid sequence of open reading frame ORF5 encoded by

15 SEQ ID NO:1669.

SEQ ID NO:1672 is the amino acid sequence of open reading frame ORF6 encoded by SEQ ID NO:1669.

SEQ ID NO:1673 is the amino acid sequence of open reading frame ORF7 encoded by SEQ ID NO:1669.

SEQ ID NO:1674 is the amino acid sequence of open reading frame ORF8 encoded by SEQ ID NO:1669.

- 5 SEQ ID NO:1675 is the amino acid sequence of open reading frame ORF9 encoded by SEQ ID NO:1669.

SEQ ID NO:1676 is the extended cDNA for contig 139 (SEQ ID NO:1467), also known as L985P.

SEQ ID NO:1677 is the L985P amino acid sequence encoded by SEQ ID NO:1676.

- 10 SEQ ID NO:1678 is the amino acid sequence of open reading frame ORF5X of SEQ ID NO:1669.

SEQ ID NO:1679 is the amino acid sequence of an open reading frame for contig 139 (SEQ ID NO:1467).

SEQ ID NO:1680-1788, set forth in the Table 9, represent cDNA clones identified by

- 15 microarray analysis of the SQL1, SCL1, SCL3 and SCL4 libraries on lung chip 5.

Table 9

SEQ ID NO:	CLONE IDENTIFIER
1680	58456
1681	58458
1682	58462
1683	58469
1684	58470
1685	58482
1686	58485
1687	58501
1688	58502
1689	58505
1690	58507
1691	58509
1692	58512
1693	58527
1694	58529
1695	58531
1696	58537
1697	58539
1698	58545
1699	59319
1700	59322
1701	59348
1702	59350
1703	59363
1704	59365
1705	59370
1706	59373
1707	59376
1708	61050
1709	61051
1710	61052
1711	61054
1712	61056
1713	61057
1714	61060
1715	61062
1716	61063
1717	61064

SEQ ID NO:	CLONE IDENTIFIER
1718	61065
1719	61066
1720	61069
1721	61070
1722	61071
1723	61074
1724	61075
1725	61077
1726	61079
1727	61080
1728	61081
1729	61083
1730	61085
1731	61086
1732	61088
1733	61090
1734	61091
1735	61093
1736	61094
1737	61096
1738	61097
1739	61099
1740	61100
1741	61103
1742	61105
1743	61106
1744	61110
1745	61113
1746	61115
1747	61117
1748	61118
1749	61119
1750	61120
1751	61122
1752	61125
1753	61126
1754	61130
1755	61133
1756	61134
1757	61135

SEQ ID NO:	CLONE IDENTIFIER
1758	61137
1759	61139
1760	61143
1761	61144
1762	61148
1763	61151
1764	61155
1765	61156
1766	61159
1767	61160
1768	61163
1769	61167
1770	61172
1771	61173
1772	61176
1773	61177
1774	61183
1775	61185
1776	61188
1777	61192
1778	61198
1779	61201
1780	61202
1781	61204
1782	61206
1783	61210
1784	61212
1785	61216
1786	61225
1787	61226
1788	61227

SEQ ID NO:1789 is the cDNA sequence of clone #47988 (L972P).

SEQ ID NO:1790 is the cDNA sequence of clone #48005 (L979P).

SEQ ID NO:1791 is an extended cDNA sequence for clone #48005 (L979P).

- 5 SEQ ID NO:1792 is an extended cDNA sequence for clone #49826 (SEQ ID NO:1279; L980P).

SEQ ID NO:1793 is an extended cDNA sequence for clone #20631 (SEQ ID NO:117; L973P).

SEQ ID NO:1794 is an extended cDNA sequence for clone #20661 (SEQ ID NO:128; L974P).

- 5 SEQ ID NO:1795 is an extended cDNA sequence for clone #50430 (SEQ ID NO:1442; L996P).

SEQ ID NO:1796 is an extended cDNA sequence for clone #26961 (SEQ ID NO:288; L977P).

- 10 SEQ ID NO:1797 is an extended cDNA sequence for clone #24928 (SEQ ID NO:1339; L978P).

SEQ ID NO:1798 is an extended cDNA sequence for clone #50507 (SEQ ID NO:1446; L984P).

SEQ ID NO:1799 is an extended cDNA sequence for clone #50645 (SEQ ID NO:1531; L988P).

- 15 SEQ ID NO:1800 is an extended cDNA sequence for clone #50628 (SEQ ID NO:1533; L1423P).

SEQ ID NO:1801 is an extended cDNA sequence for clone #50560 (SEQ ID NO:1527; L987P).

- 20 SEQ ID NO:1802 is an extended cDNA sequence for clone #27699 (SEQ ID NO:468; L998P).

SEQ ID NO:1803 is an extended cDNA sequence for clone #59303 (SEQ ID NO:949; L1425P).

SEQ ID NO:1804 is an extended cDNA sequence for clone #59314 (SEQ ID NO:1156; L1426P).

- 25 SEQ ID NO:1805 is an extended cDNA sequence for clone #59298 (SEQ ID NO:921; L1427P).

SEQ ID NO:1806 is an amino acid sequence encoded by SEQ ID NO:1791.

SEQ ID NO:1807 is an amino acid sequence encoded by SEQ ID NO:1792.

SEQ ID NO:1808 is an amino acid sequence encoded by SEQ ID NO:1793.

- SEQ ID NO:1809 is an amino acid sequence encoded by SEQ ID NO:1794.
- SEQ ID NO:1810 is an amino acid sequence encoded by SEQ ID NO:1795.
- SEQ ID NO:1811 is an amino acid sequence encoded by SEQ ID NO:1796.
- SEQ ID NO:1812 is an amino acid sequence encoded by SEQ ID NO:1797.
- 5 SEQ ID NO:1813 is an amino acid sequence encoded by SEQ ID NO:1798.
- SEQ ID NO:1814 is an amino acid sequence encoded by SEQ ID NO:1799.
- SEQ ID NO:1815 is an amino acid sequence encoded by SEQ ID NO:1800.
- SEQ ID NO:1816 is an amino acid sequence encoded by SEQ ID NO:1527 (L987P).
- SEQ ID NO:1817 is an amino acid sequence encoded by SEQ ID NO:1823.
- 10 SEQ ID NO:1818 is an amino acid sequence encoded by SEQ ID NO:1801.
- SEQ ID NO:1819 is an amino acid sequence encoded by SEQ ID NO:1802.
- SEQ ID NO:1820 is an amino acid sequence encoded by SEQ ID NO:1803.
- SEQ ID NO:1821 is an amino acid sequence encoded by SEQ ID NO:1804.
- SEQ ID NO:1822 is an amino acid sequence encoded by SEQ ID NO:1805.
- 15 SEQ ID NO:1823 is an extended cDNA sequence for clone #50560 (SEQ ID NO:1527; L987P).
- SEQ ID NO:1824 is a full length cDNA sequence for clone L872P (SEQ ID NO:34).
- SEQ ID NO:1825 is the amino acid sequence encoded by SEQ ID NO:1824.
- SEQ ID NO:1826 is the cDNA sequence encoding the N-terminal portion of L552S.
- 20 SEQ ID NO:1827-1829 are cDNA sequences of portions of L552S.
- SEQ ID NO:1830 is the N-terminal portion of L552S.
- SEQ ID NO:1831-1833 are the amino acid sequences encoded by SEQ ID NO:1827-1829, respectively.
- SEQ ID NO:1834-1856 are the amino acid sequences of peptides of L548S.
- 25 SEQ ID NO:1857-1860 are PCR primers.
- SEQ ID NO:1861 is the determined DNA sequence for a fusion of Ra12 and ORF4 of P801P.
- SEQ ID NO:1862 is the determined DNA sequence for a fusion of Ra12 and ORF5 of P801P.

- SEQ ID NO:1863 is the amino acid sequence of the fusion of Ra12 and ORF4 of P801P.
 SEQ ID NO:1864 is the amino acid sequence of the fusion of Ra12 and ORF5 of P801P.
 SEQ ID NO:1865 is the determined cDNA sequence for clone L984P_(573A).
 SEQ ID NO:1866 is the determined cDNA sequence for clone L984P_(512A).
- 5 SEQ ID NO:1867 is the determined cDNA sequence for clone L984P_(NCI-H128).
 SEQ ID NO:1868 is the determined cDNA sequence for clone L984P_(DMS-79).
 SEQ ID NO:1869 is the amino acid sequence encoded by SEQ ID NO:1865.
 SEQ ID NO:1870 is the amino acid sequence encoded by SEQ ID NO:1866.
 SEQ ID NO:1871 is the amino acid sequence encoded by SEQ ID NO:1867.
- 10 SEQ ID NO:1872 is the amino acid sequence encoded by SEQ ID NO:1868.
 SEQ ID NO:1873 is a full length cDNA sequence for clone L985P (partial sequence given in SEQ ID NO:1467).
 SEQ ID NO:1874 is the amino acid sequence for L985P encoded by SEQ ID NO:1873.
 SEQ ID NO:1875 is the predicted and determined cDNA sequence for a fusion of Ra12 and
- 15 L985P.
 SEQ ID NO:1876 is the predicted amino acid sequence of a fusion of Ra12 and L985P encoded by SEQ ID NO:1875.
 SEQ ID NO:1877 is the predicted cDNA sequence for a fusion of Ra12S and L985P.
 SEQ ID NO:1878 is the predicted amino acid sequence of a fusion of Ra12S and L985P
- 20 encoded by SEQ ID NO:1877.
 SEQ ID NO:1879 is the predicted cDNA sequence for a fusion of Ra12S and L985PEx.
 SEQ ID NO:1880 is the predicted amino acid sequence of a fusion of Ra12S and L985PEx encoded by SEQ ID NO:1879.
 SEQ ID NO:1881 is the predicted cDNA sequence the extracellular loop 2 peptide of
- 25 L985P.
 SEQ ID NO:1882 is the predicted amino acid sequence for the extracellular loop 2 peptide of L985P encoded by SEQ ID NO:1875.
 SEQ ID NO:1883 is an extended cDNA sequence for clone #59316 (SEQ ID NO:1180; L1428P).

SEQ ID NO:1884 is a first predicted amino acid sequence encoded by SEQ ID NO:1883 and designated L1428P-ORF1.

SEQ ID NO:1885 is a second predicted amino acid sequence encoded by SEQ ID NO:1883 and designated L1428P-ORF2.

- 5 SEQ ID NO:1886 is a third predicted amino acid sequence encoded by SEQ ID NO:1883 and designated L1428P-ORF3.

SEQ ID NO:1887 is a fourth predicted amino acid sequence encoded by SEQ ID NO:1883 and designated L1428P-ORF4.

- 10 SEQ ID NO:1888 is a fifth predicted amino acid sequence encoded by SEQ ID NO:1883 and designated L1428P-ORF5.

SEQ ID NO:1889 is a sixth predicted amino acid sequence encoded by SEQ ID NO:1883 and designated L1428P-ORF6.

SEQ ID NO:1890 is a seventh predicted amino acid sequence encoded by SEQ ID NO:1883 and designated L1428P-ORF7.

- 15 SEQ ID NO:1891-1900 are the nucleotide sequences for the database hits described in Table 17.

SEQ ID NO:1901-1909 are the deduced amino acid sequences encoded by the nucleotide sequences described in Table 17.

- 20 SEQ ID NO:1910 is the full-length cDNA for clone L1437P (partial sequence given in SEQ ID NO:1896).

SEQ ID NO:1911 is the forward primer PDM-433 for the coding region of clone L548S.

SEQ ID NO:1912 is the reverse primer PDM-438 for the coding region of clone L548S.

SEQ ID NO:1913 is the amino acid sequence for the expressed recombinant L548S.

SEQ ID NO:1914 is the DNA coding sequence for the recombinant L548S.

- 25 SEQ ID NO:1915 is the forward primer PDM-498 for the coding region of clone L551S

SEQ ID NO:1916 is the reverse primer PDM-499 for the coding region of clone L551S

SEQ ID NO:1917 is the amino acid sequence for the expressed recombinant L551S.

SEQ ID NO:1918 is the DNA coding sequence for the recombinant L551S.

SEQ ID NO:1919 is the forward primer PDM-479 for the coding region of clone L552S

SEQ ID NO:1920 is the reverse primer PDM-480 for the coding region of clone L552S

SEQ ID NO:1921 is the amino acid sequence for the expressed recombinant L552S.

SEQ ID NO:1922 is the DNA coding sequence for the recombinant L552S.

5 SEQ ID NO:1923 is the predicted full-length cDNA sequence for clone #19069 (partial sequence given in SEQ ID NO:90).

SEQ ID NO:1924 is the predicted full-length cDNA sequence for clone #18965 or #19002 (partial sequence given in SEQ ID NO:15).

SEQ ID NO:1925 is the deduced amino acid sequence encoded by SEQ ID NO:1923.

SEQ ID NO:1926 is the deduced amino acid sequence encoded by SEQ ID NO:1924.

10 SEQ ID NO:1927 is the determined amino acid sequence of a first L552S epitope.

SEQ ID NO:1928 is the determined amino acid sequence of a second L552S epitope.

SEQ ID NO:1929 is the determined amino acid sequence of a third L552S epitope.

SEQ ID NO:1930 is the amino acid sequence for L985P peptide #3482.

15 SEQ ID NO:1931 is an extended cDNA sequence for clone #61144 (SEQ ID NO:1761, L1439P).

SEQ ID NO:1932 is the deduced amino acid sequence encoded by SEQ ID NO:1931.

SEQ ID NO:1933 is the full-length cDNA of the NUF2R gene to which SEQ ID NO:1931 shows some sequence similarity.

SEQ ID NO:1934 is the deduced amino acid sequence encoded by SEQ ID NO:1933.

20 SEQ ID NO:1935 is a forward primer PDM-737 for the coding region of clone L552S.

SEQ ID NO:1936 is a reverse primer PDM-738 for the coding region of clone L552S.

SEQ ID NO:1937 is the amino acid sequence for the expressed recombinant L552S.

SEQ ID NO:1938 is the DNA coding sequence for the recombinant L552S.

25 SEQ ID NO:1939 is another forward primer PDM-736 for the coding region of clone L552S.

SEQ ID NO:1940 is the amino acid sequence for a second expressed recombinant L552S.

SEQ ID NO:1941 is the DNA coding sequence for a second recombinant L552S.

SEQ ID NO:1942 is the determined amino acid sequence of a fourth L552S epitope.

SEQ ID NO:1943 is the determined amino acid sequence of a first XAGE-1 epitope.

- SEQ ID NO:1944 is the determined amino acid sequence of a second XAGE-1 epitope.
- SEQ ID NO:1945 is the determined amino acid sequence of a first 20-mer peptide corresponding to amino acid residues 1-20 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1946 is the determined amino acid sequence of a second 20-mer peptide
- 5 corresponding to amino acid residues 6-25 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1947 is the determined amino acid sequence of a third 20-mer peptide corresponding to amino acid residues 11-30 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1948 is the determined amino acid sequence of a fourth 20-mer peptide corresponding to amino acid residues 16-35 of full-length L552S (SEQ ID NO:809).
- 10 SEQ ID NO:1949 is the determined amino acid sequence of a fifth 20-mer peptide corresponding to amino acid residues 21-40 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1950 is the determined amino acid sequence of a sixth 20-mer peptide corresponding to amino acid residues 26-45 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1951 is the determined amino acid sequence of a seventh 20-mer peptide
- 15 corresponding to amino acid residues 31-50 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1952 is the determined amino acid sequence of a eighth 20-mer peptide corresponding to amino acid residues 36-55 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1953 is the determined amino acid sequence of a ninth 20-mer peptide corresponding to amino acid residues 41-60 of full-length L552S (SEQ ID NO:809).
- 20 SEQ ID NO:1954 is the determined amino acid sequence of a tenth 20-mer peptide corresponding to amino acid residues 46-65 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1955 is the determined amino acid sequence of a eleventh 20-mer peptide corresponding to amino acid residues 51-70 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1955 is the determined amino acid sequence of a twelveth 20-mer peptide
- 25 corresponding to amino acid residues 56-75 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1956 is the determined amino acid sequence of a thirth 20-mer peptide corresponding to amino acid residues 61-80 of full-length L552S (SEQ ID NO:809).
- SEQ ID NO:1957 is the determined amino acid sequence of a fourteenth 20-mer peptide corresponding to amino acid residues 66-85 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1958 is the determined amino acid sequence of a fifteenth 20-mer peptide corresponding to amino acid residues 71-90 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1959 is the determined amino acid sequence of a sixteenth 20-mer peptide corresponding to amino acid residues 76-95 of full-length L552S (SEQ ID NO:809).

- 5 SEQ ID NO:1961 is the determined amino acid sequence of a seventeen 20-mer peptide corresponding to amino acid residues 81-100 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1962 is the determined amino acid sequence of a eightth 20-mer peptide corresponding to amino acid residues 86-105 of full-length L552S (SEQ ID NO:809).

- 10 SEQ ID NO:1963 is the determined amino acid sequence of a nineteenth 20-mer peptide corresponding to amino acid residues 91-110 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1964 is the determined amino acid sequence of a twentieth 20-mer peptide corresponding to amino acid residues 96-115 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1965 is the determined amino acid sequence of a twenty-first 20-mer peptide corresponding to amino acid residues 101-120 of full-length L552S (SEQ ID NO:809).

- 15 SEQ ID NO:1966 is the determined amino acid sequence of a twenty-second 20-mer peptide corresponding to amino acid residues 106-125 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1967 is the determined amino acid sequence of a twenty-third 20-mer peptide corresponding to amino acid residues 111-130 of full-length L552S (SEQ ID NO:809).

- 20 SEQ ID NO:1968 is the determined amino acid sequence of a twenty-fourth 20-mer peptide corresponding to amino acid residues 116-135 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1969 is the determined amino acid sequence of a twenty-fifth 20-mer peptide corresponding to amino acid residues 121-140 of full-length L552S (SEQ ID NO:809).

- 25 SEQ ID NO:1970 is the determined amino acid sequence of a twenty-sixth 20-mer peptide corresponding to amino acid residues 126-145 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1971 is the determined amino acid sequence of a twenth-seventh 20-mer peptide corresponding to amino acid residues 131-150 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1972 is the determined amino acid sequence of a twenty-eighth 20-mer peptide corresponding to amino acid residues 136-155 of full-length L552S (SEQ ID NO:809).

SEQ ID NO:1973 is the determined amino acid sequence of a twenty-ninth 20-mer peptide corresponding to amino acid residues 141-160 of full-length L552S (SEQ ID NO:809).

- 5 SEQ ID NO:1974 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1945.
SEQ ID NO:1975 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1946.
SEQ ID NO:1976 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1947.
SEQ ID NO:1977 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1948.
SEQ ID NO:1978 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1949.
- 10 SEQ ID NO:1979 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1950.
SEQ ID NO:1980 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1951.
SEQ ID NO:1981 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1952.
SEQ ID NO:1982 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1953.
SEQ ID NO:1983 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1954.
- 15 SEQ ID NO:1984 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1955.
SEQ ID NO:1985 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1956.
SEQ ID NO:1986 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1957.
SEQ ID NO:1987 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1958.
SEQ ID NO:1988 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1959.
- 20 SEQ ID NO:1989 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1960.
SEQ ID NO:1990 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1961.
SEQ ID NO:1991 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1962.
SEQ ID NO:1992 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1963.
SEQ ID NO:1993 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1964.
- 25 SEQ ID NO:1994 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1965.
SEQ ID NO:1995 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1966.
SEQ ID NO:1996 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1967.
SEQ ID NO:1997 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1968.
SEQ ID NO:1998 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1969.

SEQ ID NO:1999 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1970.
 SEQ ID NO:2000 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1971.
 SEQ ID NO:2001 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1972.
 SEQ ID NO:2002 is the DNA sequence which encodes the 20-mer of SEQ ID NO:1973.

5

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed generally to compositions and their use in the therapy and diagnosis of cancer, particularly lung cancer. As described further below, illustrative compositions of the present invention include, but are not restricted to, polypeptides, particularly immunogenic polypeptides, polynucleotides encoding such polypeptides, antibodies and other binding agents, antigen presenting cells (APCs) and immune system cells (*e.g.*, T cells).

The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of virology, immunology, microbiology, molecular biology and recombinant DNA techniques within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, et al. *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); Maniatis et al. *Molecular Cloning: A Laboratory Manual* (1982); DNA Cloning: A Practical Approach, vol. I & II (D. Glover, ed.); Oligonucleotide Synthesis (N. Gait, ed., 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1985); Transcription and Translation (B. Hames & S. Higgins, eds., 1984); Animal Cell Culture (R. Freshney, ed., 1986); Perbal, *A Practical Guide to Molecular Cloning* (1984).

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise.

Polypeptide Compositions

As used herein, the term "polypeptide" is used in its conventional meaning, *i.e.*, as a sequence of amino acids. The polypeptides are not limited to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide, and such terms may be used interchangeably herein unless specifically indicated otherwise. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. A polypeptide may be an entire protein, or a subsequence thereof. Particular polypeptides of interest in the context of this invention are amino acid subsequences comprising epitopes, *i.e.*, antigenic determinants substantially responsible for the immunogenic properties of a polypeptide and being capable of evoking an immune response.

Particularly illustrative polypeptides of the present invention comprise those encoded by a polynucleotide sequence set forth in any one of SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002, or a sequence that hybridizes under moderately stringent conditions, or, alternatively, under highly stringent conditions, to a polynucleotide sequence set forth in any one of SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002. Certain other illustrative polypeptides of the invention comprise amino acid sequences as set forth in any one of SEQ ID NOs:324-340, 786, 787, 789, 791, 793, 795, 797-799, 805, 806, 809, 827, 1667, 1670-1675, 1677-1679, 1806-1822, 1825, 1830-1833, 1834-1856, 1863, 1864, 1869-1872, 1874, 1876, 1878, 1880, 1882, 1884-1890, 1901-1909, 1913, 1917, 1921, 1925-1930, 1932, 1934, 1937, 1940 and 1942-1973.

The polypeptides of the present invention are sometimes herein referred to as lung tumor proteins or lung tumor polypeptides, as an indication that their identification has been based at least in part upon their increased levels of expression in lung tumor samples. Thus, a "lung tumor polypeptide" or "lung tumor protein," refers generally to a polypeptide sequence of the present invention, or a polynucleotide sequence encoding such a polypeptide, that is expressed in a substantial proportion of lung tumor samples, for example preferably greater than about 20%, more preferably greater than about 30%, and most preferably greater than about 50% or more of lung tumor samples tested, at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in normal tissues, as determined using a representative assay provided herein. A lung tumor polypeptide sequence of the invention, based upon its increased level of expression in tumor cells, has particular utility both as a diagnostic marker as well as a therapeutic target, as further described below.

In certain preferred embodiments, the polypeptides of the invention are immunogenic, *i.e.*, they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with lung cancer. Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such screens can be performed using methods such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ¹²⁵I-labeled Protein A.

As would be recognized by the skilled artisan, immunogenic portions of the polypeptides disclosed herein are also encompassed by the present invention. An "immunogenic portion," as used herein, is a fragment of an immunogenic polypeptide of the invention that itself is immunologically reactive (*i.e.*, specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such as those

summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen
 5 (i.e., they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well-known techniques.

In one preferred embodiment, an immunogenic portion of a polypeptide of the present invention is a portion that reacts with antisera and/or T-cells at a level that is not
 10 substantially less than the reactivity of the full-length polypeptide (e.g., in an ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic portion is at least about 50%, preferably at least about 70% and most preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic
 15 activity greater than that of the corresponding full-length polypeptide, e.g., having greater than about 100% or 150% or more immunogenic activity.

In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other illustrative immunogenic portions will contain a small N- and/or
 20 C-terminal deletion (e.g., 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

In another embodiment, a polypeptide composition of the invention may also comprise one or more polypeptides that are immunologically reactive with T cells and/or antibodies generated against a polypeptide of the invention, particularly a
 25 polypeptide having an amino acid sequence disclosed herein, or to an immunogenic fragment or variant thereof.

In another embodiment of the invention, polypeptides are provided that comprise one or more polypeptides that are capable of eliciting T cells and/or antibodies that are immunologically reactive with one or more polypeptides described herein, or one

or more polypeptides encoded by contiguous nucleic acid sequences contained in the polynucleotide sequences disclosed herein, or immunogenic fragments or variants thereof, or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency.

5 The present invention, in another aspect, provides polypeptide fragments comprising at least about 5, 10, 15, 20, 25, 50, or 100 contiguous amino acids, or more, including all intermediate lengths, of a polypeptide compositions set forth herein, such as those set forth in SEQ ID NOs:324-340, 786, 787, 789, 791, 793, 795, 797-799, 805, 806, 809, 827, 1667, 1670-1675, 1677-1679, 1806-1822, 1825, 1830-1833, 1834-1856, 1863,
10 1864, 1869-1872, 1874, 1876, 1878, 1880, 1882, 1884-1890, 1901-1909, 1913, 1917, 1921, 1925-1930, 1932, 1934, 1937, 1940 and 1942-1973, or those encoded by a polynucleotide sequence set forth in a sequence of SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883,
15 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002.

 In another aspect, the present invention provides variants of the polypeptide compositions described herein. Polypeptide variants generally encompassed by the present invention will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described below),
20 along its length, to a polypeptide sequences set forth herein.

 In one preferred embodiment, the polypeptide fragments and variants provide by the present invention are immunologically reactive with an antibody and/or T-cell that reacts with a full-length polypeptide specifically set for the herein.

 In another preferred embodiment, the polypeptide fragments and variants
25 provided by the present invention exhibit a level of immunogenic activity of at least about 50%, preferably at least about 70%, and most preferably at least about 90% or more of that exhibited by a full-length polypeptide sequence specifically set forth herein.

 A polypeptide "variant," as the term is used herein, is a polypeptide that typically differs from a polypeptide specifically disclosed herein in one or more

substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more of the above polypeptide sequences of the invention and evaluating their immunogenic activity as described herein and/or using any of a number of techniques well known in the art.

5 For example, certain illustrative variants of the polypeptides of the invention include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other illustrative variants include variants in which a small portion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

10 In many instances, a variant will contain conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. As described above, modifications may be made in the structure of the
 15 polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics, *e.g.*, with immunogenic characteristics. When it is desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, immunogenic variant or portion of a polypeptide of the invention, one skilled in the art will typically change one or
 20 more of the codons of the encoding DNA sequence according to Table 10.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines
 25 that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed compositions, or

corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

TABLE 10

5

Amino Acids			Codons						
Alanine	Ala	A	GCA	GCC	GCG	GCU			
Cysteine	Cys	C	UGC	UGU					
Aspartic acid	Asp	D	GAC	GAU					
Glutamic acid	Glu	E	GAA	GAG					
Phenylalanine	Phe	F	UUC	UUU					
Glycine	Gly	G	GGA	GGC	GGG	GGU			
Histidine	His	H	CAC	CAU					
Isoleucine	Ile	I	AUA	AUC	AUU				
Lysine	Lys	K	AAA	AAG					
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU	
Methionine	Met	M	AUG						
Asparagine	Asn	N	AAC	AAU					
Proline	Pro	P	CCA	CCC	CCG	CCU			
Glutamine	Gln	Q	CAA	CAG					
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU	
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU	
Threonine	Thr	T	ACA	ACC	ACG	ACU			
Valine	Val	V	GUA	GUC	GUG	GUU			
Tryptophan	Trp	W	UGG						
Tyrosine	Tyr	Y	UAC	UAU					

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, 10 incorporated herein by reference). It is accepted that the relative hydropathic character of

the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics

5 (Kyte and Doolittle, 1982). These values are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (−0.4); threonine (−0.7); serine (−0.8); tryptophan (−0.9); tyrosine (−1.3); proline (−1.6); histidine (−3.2); glutamate (−3.5); glutamine (−3.5); aspartate (−3.5); asparagine (−3.5); lysine (−3.9); and arginine (−4.5).

10 It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even

15 more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

20 As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 \pm 1); glutamate (+3.0 \pm 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (−0.4); proline (−0.5 \pm 1); alanine (−0.5); histidine (−0.5); cysteine (−1.0); methionine (−1.3); valine (−1.5); leucine (−1.8); isoleucine (−1.8); tyrosine (−2.3);

25 phenylalanine (−2.5); tryptophan (−3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those

within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl-, methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

Amino acid substitutions may further be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydrophobic nature of the polypeptide.

As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-
 5 His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

When comparing polypeptide sequences, two sequences are said to be “identical” if the sequence of amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences
 10 are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A “comparison window” as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the
 15 Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff,
 20 M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971)
 25 *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment.

In one preferred approach, the “percentage of sequence identity” is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical amino acid residue occurs in both sequences to yield the number of matched positions,

dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Within other illustrative embodiments, a polypeptide may be a fusion polypeptide that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the polypeptide or to enable the polypeptide to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the polypeptide.

Fusion polypeptides may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion polypeptide is expressed as a recombinant polypeptide, allowing the production of increased levels, relative to a non-fused polypeptide, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion polypeptide that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a

flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

The fusion polypeptide can comprise a polypeptide as described herein together with an unrelated immunogenic protein, such as an immunogenic protein capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see*, for example, Stoute et al. *New Engl. J. Med.*, 336:86-91, 1997).

In one preferred embodiment, the immunological fusion partner is derived from a *Mycobacterium* sp., such as a *Mycobacterium tuberculosis*-derived Ra12 fragment. Ra12 compositions and methods for their use in enhancing the expression and/or immunogenicity of heterologous polynucleotide/polypeptide sequences is described in U.S. Patent Application 60/158,585, the disclosure of which is incorporated herein by reference in its entirety. Briefly, Ra12 refers to a polynucleotide region that is a subsequence of a *Mycobacterium tuberculosis* MTB32A nucleic acid. MTB32A is a serine protease of 32 KD molecular weight encoded by a gene in virulent and avirulent strains of *M. tuberculosis*. The nucleotide sequence and amino acid sequence of MTB32A have been

described (for example, U.S. Patent Application 60/158,585; see also, Skeiky *et al.*, *Infection and Immun.* (1999) 67:3998-4007, incorporated herein by reference). Surprisingly, it was discovered that a 14KD C-terminal fragment of the MTB32A coding sequence expresses at high levels on its own and remains as a soluble polypeptide throughout the purification process. Moreover, this fragment may enhance the immunogenicity of heterologous antigenic polypeptides with which it is fused. This 14 KD C-terminal fragment is referred to herein as Ra12 and represents a fragment comprising some or all of amino acid residues 192 to 323 of MTB32A.

Other preferred Ra12 polynucleotides generally comprise at least about 15 consecutive nucleotides, at least about 30 nucleotides, at least about 60 nucleotides, at least about 100 nucleotides, at least about 200 nucleotides, or at least about 300 nucleotides that encode a portion of a Ra12 polypeptide.

Ra12 polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a Ra12 polypeptide or a portion thereof) or may comprise a variant of such a sequence. Ra12 polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion polypeptide is not substantially diminished, relative to a fusion polypeptide comprising a native Ra12 polypeptide. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native Ra12 polypeptide or a portion thereof.

Within other preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza B* (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural

protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the *LytA* gene; *Gene* 43:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (see *Biotechnology* 10:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion polypeptide. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

Yet another illustrative embodiment involves fusion polypeptides, and the polynucleotides encoding them, wherein the fusion partner comprises a targeting signal capable of directing a polypeptide to the endosomal/lysosomal compartment, as described in U.S. Patent No. 5,633,234. An immunogenic polypeptide of the invention, when fused with this targeting signal, will associate more efficiently with MHC class II molecules and thereby provide enhanced in vivo stimulation of CD4⁺ T-cells specific for the polypeptide.

Polypeptides of the invention are prepared using any of a variety of well known synthetic and/or recombinant techniques, the latter of which are further described below. Polypeptides, portions and other variants generally less than about 150 amino acids can be generated by synthetic means, using techniques well known to those of ordinary skill in the art. In one illustrative example, such polypeptides are synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of

polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

In general, polypeptide compositions (including fusion polypeptides) of the invention are isolated. An "isolated" polypeptide is one that is removed from its original environment. For example, a naturally-occurring protein or polypeptide is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are also purified, *e.g.*, are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure.

10 Polynucleotide Compositions

The present invention, in other aspects, provides polynucleotide compositions. The terms "DNA" and "polynucleotide" are used essentially interchangeably herein to refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. "Isolated," as used herein, means that a polynucleotide is substantially
15 away from other coding sequences, and that the DNA molecule does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

20 As will be understood by those skilled in the art, the polynucleotide compositions of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

25 As will be also recognized by the skilled artisan, polynucleotides of the invention may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules may include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-

one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

5 Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a polypeptide/protein of the invention or a portion thereof) or may comprise a sequence that encodes a variant or derivative, preferably and immunogenic variant or derivative, of such a sequence.

Therefore, according to another aspect of the present invention,
 10 polynucleotide compositions are provided that comprise some or all of a polynucleotide sequence set forth in any one of SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002, complements of a
 15 polynucleotide sequence set forth in any one of SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002, and degenerate variants of a polynucleotide sequence set forth in any one of SEQ ID NOs:1-57, 59-323,
 20 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and 1974-2002. In certain preferred embodiments, the polynucleotide sequences set forth herein encode immunogenic polypeptides, as described above.

25 In other related embodiments, the present invention provides polynucleotide variants having substantial identity to the sequences disclosed herein in SEQ ID NOs:1-57, 59-323, 341-782, 784-785, 788, 790, 792, 794, 796, 800-804, 807, 808, 810-826, 828-1664, 1668, 1669, 1676, 1680-1805, 1824, 1826-1829, 1865-1868, 1873, 1875, 1877, 1879, 1881, 1883, 1891-1900, 1910, 1914, 1918, 1922-1924, 1931, 1933, 1938, 1941 and

1974-2002, for example those comprising at least 70% sequence identity, preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide sequence of this invention using the methods described herein, (*e.g.*, BLAST analysis using standard parameters, as described below). One skilled
 5 in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

Typically, polynucleotide variants will contain one or more substitutions, additions, deletions and/or insertions, preferably such that the immunogenicity of the
 10 polypeptide encoded by the variant polynucleotide is not substantially diminished relative to a polypeptide encoded by a polynucleotide sequence specifically set forth herein). The term "variants" should also be understood to encompass homologous genes of xenogenic origin.

In additional embodiments, the present invention provides polynucleotide
 15 fragments comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 10, 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It
 20 will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like.

In another embodiment of the invention, polynucleotide compositions are
 25 provided that are capable of hybridizing under moderate to high stringency conditions to a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include

prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-60°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS. One skilled in the art will understand that the stringency of hybridization can be readily manipulated, such as by
 5 altering the salt content of the hybridization solution and/or the temperature at which the hybridization is performed. For example, in another embodiment, suitable highly stringent hybridization conditions include those described above, with the exception that the temperature of hybridization is increased, *e.g.*, to 60-65°C or 65-70°C.

In certain preferred embodiments, the polynucleotides described above, *e.g.*,
 10 polynucleotide variants, fragments and hybridizing sequences, encode polypeptides that are immunologically cross-reactive with a polypeptide sequence specifically set forth herein. In other preferred embodiments, such polynucleotides encode polypeptides that have a level of immunogenic activity of at least about 50%, preferably at least about 70%, and more preferably at least about 90% of that for a polypeptide sequence specifically set forth
 15 herein.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length
 20 may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative polynucleotide segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the
 25 like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

When comparing polynucleotide sequences, two sequences are said to be “identical” if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are

typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A “comparison window” as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same
 5 number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of
 10 evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS*
 15 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

20 Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP,
 25 BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402

and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to

the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

Therefore, in another embodiment of the invention, a mutagenesis approach, such as site-specific mutagenesis, is employed for the preparation of immunogenic variants and/or derivatives of the polypeptides described herein. By this approach, specific modifications in a polypeptide sequence can be made through mutagenesis of the underlying polynucleotides that encode them. These techniques provides a straightforward approach to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the polynucleotide.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Mutations may be employed in a selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise change the properties of the polynucleotide itself, and/or alter the properties, activity, composition, stability, or primary sequence of the encoded polypeptide.

In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or more properties of the encoded polypeptide, such as the immunogenicity of a polypeptide vaccine. The techniques of site-specific mutagenesis are well-known in the art, and are widely used to create variants of both polypeptides and polynucleotides. For example, site-specific

mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25 nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of the sequence being altered.

5 As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also
10 routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide.
15 An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears
20 the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis provides a means of producing potentially useful
25 species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding these methods and protocols are found in the teachings of Maloy *et al.*, 1994; Segal, 1976;

Prokop and Bajpai, 1991; Kuby, 1994; and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term “oligonucleotide directed mutagenesis procedure” refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term “oligonucleotide directed mutagenesis procedure” is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U. S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

In another approach for the production of polypeptide variants of the present invention, recursive sequence recombination, as described in U.S. Patent No. 5,837,458, may be employed. In this approach, iterative cycles of recombination and screening or selection are performed to “evolve” individual polynucleotide variants of the invention having, for example, enhanced immunogenic activity.

In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long contiguous sequence disclosed herein will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all

intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned, such as the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Polynucleotide molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide sequence disclosed herein, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

The use of a hybridization probe of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequences set forth herein, or to any continuous portion of the sequences, from about 15-25 nucleotides in length up to

and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™ technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

The nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of

increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

5 According to another embodiment of the present invention, polynucleotide compositions comprising antisense oligonucleotides are provided. Antisense oligonucleotides have been demonstrated to be effective and targeted inhibitors of protein synthesis, and, consequently, provide a therapeutic approach by which a disease can be treated by inhibiting the synthesis of proteins that contribute to the disease. The efficacy of
10 antisense oligonucleotides for inhibiting protein synthesis is well established. For example, the synthesis of polygalacturonase and the muscarine type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829). Further, examples of antisense inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene
15 (MDG1), ICAM-1, E-selectin, STK-1, striatal GABA_A receptor and human EGF (Jaskulski *et al.*, Science. 1988 Jun 10;240(4858):1544-6; Vasanthakumar and Ahmed, Cancer Commun. 1989;1(4):225-32; Peris *et al.*, Brain Res Mol Brain Res. 1998 Jun 15;57(2):310-20; U. S. Patent 5,801,154; U.S. Patent 5,789,573; U. S. Patent 5,718,709 and U.S. Patent 5,610,288). Antisense constructs have also been described that inhibit and can be used to
20 treat a variety of abnormal cellular proliferations, *e.g.* cancer (U. S. Patent 5,747,470; U. S. Patent 5,591,317 and U. S. Patent 5,783,683).

Therefore, in certain embodiments, the present invention provides oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof.
25 In one embodiment, the antisense oligonucleotides comprise DNA or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred

compositions comprise a sequence region that is complementary, and more preferably substantially-complementary, and even more preferably, completely complementary to one or more portions of polynucleotides disclosed herein. Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence and
 5 determination of secondary structure, T_m , binding energy, and relative stability. Antisense compositions may be selected based upon their relative inability to form dimers, hairpins, or other secondary structures that would reduce or prohibit specific binding to the target mRNA in a host cell. Highly preferred target regions of the mRNA, are those which are at or near the AUG translation initiation codon, and those sequences which are substantially
 10 complementary to 5' regions of the mRNA. These secondary structure analyses and target site selection considerations can be performed, for example, using v.4 of the OLIGO primer analysis software and/or the BLASTN 2.0.5 algorithm software (Altschul *et al.*, Nucleic Acids Res. 1997, 25(17):3389-402).

The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a
 15 hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic domain from the nuclear localization sequence of SV40 T-antigen (Morris *et al.*, Nucleic Acids Res. 1997 Jul 15;25(14):2730-6). It has been demonstrated that several molecules of the MPG peptide coat the antisense oligonucleotides and can be delivered into cultured
 20 mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the oligonucleotide to nuclease and the ability to cross the plasma membrane.

According to another embodiment of the invention, the polynucleotide compositions described herein are used in the design and preparation of ribozyme
 25 molecules for inhibiting expression of the tumor polypeptides and proteins of the present invention in tumor cells. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, Proc Natl Acad Sci U S A. 1987 Dec;84(24):8788-92; Forster and Symons, Cell. 1987 Apr 24;49(2):211-20). For example, a large number of

ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, Cell. 1981 Dec;27(3 Pt 2):487-96; Michel and Westhof, J Mol Biol. 1990 Dec 5;216(3):585-610; Reinhold-Hurek and Shub, Nature. 1992 May 14;357(6374):173-6).

- 5 This specificity has been attributed to the requirement that the substrate bind via specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and
 10 thus can cleave other RNA molecules) under physiological conditions. In general, enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-
 15 pairing, and once bound to the correct site, acts enzymatically to cut the target RNA. Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

20 The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme
 25 molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense

molecules do not prevent their action (Woolf *et al.*, Proc Natl Acad Sci U S A. 1992 Aug 15;89(16):7305-9). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis δ virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi *et al.* Nucleic Acids Res. 1992 Sep 11;20(17):4559-65. Examples of hairpin motifs are described by Hampel *et al.* (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz, Biochemistry 1989 Jun 13;28(12):4929-33; Hampel *et al.*, Nucleic
 5 Acids Res. 1990 Jan 25;18(2):299-304 and U. S. Patent 5,631,359. An example of the hepatitis δ virus motif is described by Perrotta and Been, Biochemistry. 1992 Dec 1;31(47):11843-52; an example of the RNaseP motif is described by Guerrier-Takada *et al.*, Cell. 1983 Dec;35(3 Pt 2):849-57; Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, Cell. 1990 May 18;61(4):685-96; Saville and Collins, Proc
 10 Natl Acad Sci U S A. 1991 Oct 1;88(19):8826-30; Collins and Olive, Biochemistry. 1993 Mar 23;32(11):2795-9); and an example of the Group I intron is described in (U. S. Patent 4,987,071). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that
 15 substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such
 20 ribozymes can also be optimized for delivery. While specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their

degradation by serum ribonucleases (see *e.g.*, Int. Pat. Appl. Publ. No. WO 92/07065; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; Eur. Pat. Appl. Publ. No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided in Int. Pat. Appl. Publ. No. WO 94/02595 and Int. Pat. Appl. Publ. No. WO 93/23569, each specifically incorporated herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase

enzyme is expressed in the appropriate cells. Ribozymes expressed from such promoters have been shown to function in mammalian cells. Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-
 5 associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

In another embodiment of the invention, peptide nucleic acids (PNAs) compositions are provided. PNA is a DNA mimic in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, *Antisense Nucleic Acid Drug Dev.* 1997
 10 7(4) 431-37). PNA is able to be utilized in a number methods that traditionally have used RNA or DNA. Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (*Trends Biotechnol* 1997 Jun;15(6):224-9). As such, in certain embodiments, one
 15 may prepare PNA sequences that are complementary to one or more portions of the ACE mRNA sequence, and such PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal
 20 phosphodiester backbone of DNA (Nielsen *et al.*, *Science* 1991 Dec 6;254(5037):1497-500; Hanvey *et al.*, *Science*. 1992 Nov 27;258(5087):1481-5; Hyrup and Nielsen, *Bioorg Med Chem.* 1996 Jan;4(1):5-23). This chemistry has three important consequences: firstly, in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the need to develop a stereoselective synthesis;
 25 and thirdly, PNA synthesis uses standard Boc or Fmoc protocols for solid-phase peptide synthesis, although other methods, including a modified Merrifield method, have been used.

PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc

protocols are straightforward using manual or automated protocols (Norton *et al.*, Bioorg Med Chem. 1995 Apr;3(4):437-45). The manual protocol lends itself to the production of chemically modified PNAs or the simultaneous synthesis of families of closely related PNAs.

5 As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is suggested that, in producing PNAs with adjacent purines, one should repeat the coupling of
10 residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography, providing yields and purity of product similar to those observed during the synthesis of peptides.

Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain
15 a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or for specific functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications
20 of PNAs (for example, Norton *et al.*, Bioorg Med Chem. 1995 Apr;3(4):437-45; Petersen *et al.*, J Pept Sci. 1995 May-Jun;1(3):175-83; Orum *et al.*, Biotechniques. 1995 Sep;19(3):472-80; Footer *et al.*, Biochemistry. 1996 Aug 20;35(33):10673-9; Griffith *et al.*, Nucleic Acids Res. 1995 Aug 11;23(15):3003-8; Pardridge *et al.*, Proc Natl Acad Sci U S A. 1995 Jun 6;92(12):5592-6; Boffa *et al.*, Proc Natl Acad Sci U S A. 1995 Mar
25 14;92(6):1901-5; Gambacorti-Passerini *et al.*, Blood. 1996 Aug 15;88(4):1411-7; Armitage *et al.*, Proc Natl Acad Sci U S A. 1997 Nov 11;94(23):12320-5; Seeger *et al.*, Biotechniques. 1997 Sep;23(3):512-7). U.S. Patent No. 5,700,922 discusses PNA-DNA-PNA chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.

Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (Anal Chem. 1993 Dec 15;65(24):3545-9) and Jensen *et al.* (Biochemistry. 1997 Apr 22;36(16):5072-7). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative
 5 binding kinetics and stoichiometry. Similar types of measurements were made by Jensen *et al.* using BIAcore™ technology.

Other applications of PNAs that have been described and will be apparent to the skilled artisan include use in DNA strand invasion, antisense inhibition, mutational analysis, enhancers of transcription, nucleic acid purification, isolation of transcriptionally
 10 active genes, blocking of transcription factor binding, genome cleavage, biosensors, *in situ* hybridization, and the like.

Polynucleotide Identification, Characterization and Expression

Polynucleotides compositions of the present invention may be identified, prepared and/or manipulated using any of a variety of well established techniques (see
 15 generally, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989, and other like references). For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay
 20 provided herein). Such screens may be performed, for example, using the microarray technology of Affymetrix, Inc. (Santa Clara, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller et al., *Proc. Natl. Acad. Sci. USA* 94:2150-2155, 1997). Alternatively, polynucleotides may be amplified from cDNA prepared from cells
 25 expressing the proteins described herein, such as tumor cells.

Many template dependent processes are available to amplify a target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR™) which is described in detail in U.S. Patent Nos.

4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCRTM, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA
 5 polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product
 10 and the process is repeated. Preferably reverse transcription and PCRTM amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Any of a number of other template dependent processes, many of which are variations of the PCRTM amplification technique, are readily known and available in the
 15 art. Illustratively, some such methods include the ligase chain reaction (referred to as LCR), described, for example, in Eur. Pat. Appl. Publ. No. 320,308 and U.S. Patent No. 4,883,750; Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880; Strand Displacement Amplification (SDA) and Repair Chain Reaction (RCR). Still other amplification methods are described in Great Britain Pat. Appl. No. 2 202 328, and in PCT
 20 Intl. Pat. Appl. Publ. No. PCT/US89/01025. Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. Eur. Pat. Appl. Publ. No. 329,822 describes a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded
 25 DNA (dsDNA). PCT Intl. Pat. Appl. Publ. No. WO 89/06700 describes a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. Other amplification methods such as "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) are also well-known to those of skill in the art.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.*, a tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (*e.g.*, by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences can then be assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, amplification techniques, such as those described above, can be useful for obtaining a full length coding sequence from a partial cDNA sequence. One such amplification technique is inverse PCR (*see* Triglia et al., *Nucl. Acids Res.* 16:8186, 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified

sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid
 5 amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Applic. 1*:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids. Res. 19*:3055-60, 1991). Other methods employing
 10 amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.*, NCBI BLAST searches), and such ESTs may be used to
 15 generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct
 20 expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in
 25 some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA
 5 shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

10 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site
 15 located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.*
 20 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) *Science* 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

25 A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure).

Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements.

These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y.

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the

PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, any of a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone

- or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311. Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl.*
- 5 *Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).
- 10 An insect system may also be used to express a polypeptide of interest. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding the polypeptide may be cloned into a non-
- 15 essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed (Engelhard, E. K. et al. (1994) *Proc. Natl. Acad. Sci.* 91 :3224-3227).
- 20 In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to
- 25 obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be
 5 needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may
 10 be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the
 15 expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, COS, HeLa,
 20 MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a
 25 polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection,

and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell
 5 lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1990) *Cell* 22:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al.
 10 (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or
 15 hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-51). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression
 20 attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence,
 25 recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include, for example, membrane,
 5 solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent
 10 assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological
 15 Methods, a Laboratory Manual, APS Press, St Paul. Minn.) and Maddox, D. E. et al. (1983; *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to
 20 polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides.
 25 These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* 3:263-281) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; *DNA Cell Biol.* 12:441-453).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) *J. Am. Chem. Soc.* 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin

Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Antibody Compositions, Fragments Thereof and Other Binding Agents

According to another aspect, the present invention further provides binding
 5 agents, such as antibodies and antigen-binding fragments thereof, that exhibit immunological binding to a tumor polypeptide disclosed herein, or to a portion, variant or derivative thereof. An antibody, or antigen-binding fragment thereof, is said to "specifically bind," "immunologically bind," and/or is "immunologically reactive" to a polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA
 10 assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions.

Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of
 15 immunological binding interactions can be expressed in terms of the dissociation constant (K_d) of the interaction, wherein a smaller K_d represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates depend on the concentrations of
 20 the complex partners, the affinity of the interaction, and on geometric parameters that equally influence the rate in both directions. Thus, both the "on rate constant" (K_{on}) and the "off rate constant" (K_{off}) can be determined by calculation of the concentrations and the actual rates of association and dissociation. The ratio of K_{off}/K_{on} enables cancellation of all parameters not related to affinity, and is thus equal to the dissociation constant K_d . See,
 25 generally, Davies et al. (1990) Annual Rev. Biochem. 59:439-473.

An "antigen-binding site," or "binding portion" of an antibody refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of

the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains are referred to as "hypervariable regions" which are interposed between more conserved flanking stretches known as "framework regions," or "FRs". Thus the term "FR" refers to amino acid sequences which are naturally found
 5 between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each
 10 of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs."

Binding agents may be further capable of differentiating between patients with and without a cancer, such as lung cancer, using the representative assays provided herein. For example, antibodies or other binding agents that bind to a tumor protein will
 15 preferably generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, more preferably at least about 30% of patients. Alternatively, or in addition, the antibody will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.*, blood, sera, sputum, urine
 20 and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. Preferably, a statistically significant number of samples with and without the disease will be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may
 25 be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of

techniques known to those of ordinary skill in the art. *See, e.g.,* Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into

5 suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior

10 immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example,

15 affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired

20 specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may

25 be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single

colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

A number of therapeutically useful molecules are known in the art which comprise antigen-binding sites that are capable of exhibiting immunological binding properties of an antibody molecule. The proteolytic enzyme papain preferentially cleaves IgG molecules to yield several fragments, two of which (the "F(ab)" fragments) each comprise a covalent heterodimer that includes an intact antigen-binding site. The enzyme pepsin is able to cleave IgG molecules to provide several fragments, including the "F(ab')₂" fragment which comprises both antigen-binding sites. An "Fv" fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly derived using recombinant techniques known in the art. The Fv fragment includes a non-covalent V_H::V_L heterodimer including an antigen-binding site which retains much of the antigen recognition and binding capabilities of the native antibody molecule. Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976) Biochem 15:2706-2710; and Ehrlich et al. (1980) Biochem 19:4091-4096.

A single chain Fv ("sFv") polypeptide is a covalently linked V_H::V_L heterodimer which is expressed from a gene fusion including V_H- and V_L-encoding genes linked by a peptide-encoding linker. Huston et al. (1988) Proc. Nat. Acad. Sci. USA 85(16):5879-5883. A number of methods have been described to discern chemical structures for converting the naturally aggregated--but chemically separated--light and

heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, *e.g.*, U.S. Pat. Nos. 5,091,513 and 5,132,405, to Huston et al.; and U.S. Pat. No. 4,946,778, to Ladner et al.

5 Each of the above-described molecules includes a heavy chain and a light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRS and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light
10 chain, these regions are denoted as "CDR1," "CDR2," and "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (*e.g.*, a CDR1, CDR2 or CDR3) is referred to herein as a "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid
15 residues of CDRs form extensive contact with bound antigen, wherein the most extensive antigen contact is with the heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region. Some
20 FR residues may contact bound antigen; however, FRs are primarily responsible for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRS. Within FRs, certain amino residues and certain structural features are very highly conserved. In this regard, all V region sequences contain an internal disulfide loop of around 90 amino acid residues. When the V regions fold into a binding-site, the CDRs are
25 displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of FRs which influence the folded shape of the CDR loops into certain "canonical" structures--regardless of the precise CDR amino acid sequence. Further, certain FR residues are known to participate in non-covalent interdomain contacts which stabilize the interaction of the antibody heavy and light chains.

A number of "humanized" antibody molecules comprising an antigen-binding site derived from a non-human immunoglobulin have been described, including chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter et al. (1991) Nature 349:293-299; Lobuglio et al. (1989) Proc. Nat. Acad. Sci. USA 86:4220-4224; Shaw et al. (1987) J Immunol. 138:4534-4538; and Brown et al. (1987) Cancer Res. 47:3577-3583), rodent CDRs grafted into a human supporting FR prior to fusion with an appropriate human antibody constant domain (Riechmann et al. (1988) Nature 332:323-327; Verhoeyen et al. (1988) Science 239:1534-1536; and Jones et al. (1986) Nature 321:522-525), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These "humanized" molecules are designed to minimize unwanted immunological response toward rodent antihuman antibody molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients.

As used herein, the terms "veneered FRs" and "recombinantly veneered FRs" refer to the selective replacement of FR residues from, *e.g.*, a rodent heavy or light chain V region, with human FR residues in order to provide a xenogeneic molecule comprising an antigen-binding site which retains substantially all of the native FR polypeptide folding structure. Veneering techniques are based on the understanding that the ligand binding characteristics of an antigen-binding site are determined primarily by the structure and relative disposition of the heavy and light chain CDR sets within the antigen-binding surface. Davies et al. (1990) Ann. Rev. Biochem. 59:439-473. Thus, antigen binding specificity can be preserved in a humanized antibody only wherein the CDR structures, their interaction with each other, and their interaction with the rest of the V region domains are carefully maintained. By using veneering techniques, exterior (*e.g.*, solvent-accessible) FR residues which are readily encountered by the immune system are selectively replaced with human residues to provide a hybrid molecule that comprises either a weakly immunogenic, or substantially non-immunogenic veneered surface.

The process of veneering makes use of the available sequence data for human antibody variable domains compiled by Kabat et al., in Sequences of Proteins of Immunological Interest, 4th ed., (U.S. Dept. of Health and Human Services, U.S. Government Printing Office, 1987), updates to the Kabat database, and other accessible
 5 U.S. and foreign databases (both nucleic acid and protein). Solvent accessibilities of V region amino acids can be deduced from the known three-dimensional structure for human and murine antibody fragments. There are two general steps in veneering a murine antigen-binding site. Initially, the FRs of the variable domains of an antibody molecule of interest are compared with corresponding FR sequences of human variable domains obtained from
 10 the above-identified sources. The most homologous human V regions are then compared residue by residue to corresponding murine amino acids. The residues in the murine FR which differ from the human counterpart are replaced by the residues present in the human moiety using recombinant techniques well known in the art. Residue switching is only carried out with moieties which are at least partially exposed (solvent accessible), and care
 15 is exercised in the replacement of amino acid residues which may have a significant effect on the tertiary structure of V region domains, such as proline, glycine and charged amino acids.

In this manner, the resultant "veneered" murine antigen-binding sites are thus designed to retain the murine CDR residues, the residues substantially adjacent to the
 20 CDRs, the residues identified as buried or mostly buried (solvent inaccessible), the residues believed to participate in non-covalent (*e.g.*, electrostatic and hydrophobic) contacts between heavy and light chain domains, and the residues from conserved structural regions of the FRs which are believed to influence the "canonical" tertiary structures of the CDR loops. These design criteria are then used to prepare recombinant nucleotide sequences
 25 which combine the CDRs of both the heavy and light chain of a murine antigen-binding site into human-appearing FRs that can be used to transfect mammalian cells for the expression of recombinant human antibodies which exhibit the antigen specificity of the murine antibody molecule.

In another embodiment of the invention, monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ^{90}Y , ^{123}I , ^{125}I , ^{131}I , ^{186}Re , ^{188}Re , ^{211}At , and ^{212}Bi .

5 Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diphtheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable
10 monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a
15 halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody,
20 and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the
25 catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

T Cell Compositions

The present invention, in another aspect, provides T cells specific for a tumor polypeptide disclosed herein, or for a variant or derivative thereof. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells
 5 may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the Isolex™ System, available from Nexell Therapeutics, Inc. (Irvine, CA; see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-
 10 human mammals, cell lines or cultures.

T cells may be stimulated with a polypeptide, polynucleotide encoding a polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide of interest. Preferably, a tumor
 15 polypeptide or polynucleotide of the invention is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a polypeptide of the present invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may
 20 be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen et al., *Cancer Res.* 54:1065-1070, 1994. Alternatively, detection of the proliferation of T cells
 25 may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.*, by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a tumor polypeptide (100 ng/ml - 100 µg/ml, preferably 200 ng/ml - 25 µg/ml) for 3 - 7 days will typically result in at least a two fold

increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine release (*e.g.*, TNF or IFN- γ) is indicative of T cell activation (*see* Coligan et al., Current Protocols in Immunology, vol. 1, Wiley Interscience
 5 (Greene 1998)). T cells that have been activated in response to a tumor polypeptide, polynucleotide or polypeptide-expressing APC may be CD4⁺ and/or CD8⁺. Tumor polypeptide-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

10 For therapeutic purposes, CD4⁺ or CD8⁺ T cells that proliferate in response to a tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the
 15 addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence of the tumor polypeptide can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

20 T Cell Receptor Compositions

The T cell receptor (TCR) consists of 2 different, highly variable polypeptide chains, termed the T-cell receptor α and β chains, that are linked by a disulfide bond (Janeway, Travers, Walport. *Immunobiology*. Fourth Ed., 148-159. Elsevier Science Ltd/Garland Publishing. 1999). The α/β heterodimer complexes with the invariant CD3
 25 chains at the cell membrane. This complex recognizes specific antigenic peptides bound to MHC molecules. The enormous diversity of TCR specificities is generated much like immunoglobulin diversity, through somatic gene rearrangement. The β chain genes contain over 50 variable (V), 2 diversity (D), over 10 joining (J) segments, and 2 constant region

segments (C). The α chain genes contain over 70 V segments, and over 60 J segments but no D segments, as well as one C segment. During T cell development in the thymus, the D to J gene rearrangement of the β chain occurs, followed by the V gene segment rearrangement to the DJ. This functional VDJ β exon is transcribed and spliced to join to a C β . For the α chain, a V α gene segment rearranges to a J α gene segment to create the functional exon that is then transcribed and spliced to the C α . Diversity is further increased during the recombination process by the random addition of P and N-nucleotides between the V, D, and J segments of the β chain and between the V and J segments in the α chain (Janeway, Travers, Walport. *Immunobiology*. Fourth Ed., 98 and 150. Elsevier Science Ltd/Garland Publishing. 1999).

The present invention, in another aspect, provides TCRs specific for a polypeptide disclosed herein, or for a variant or derivative thereof. In accordance with the present invention, polynucleotide and amino acid sequences are provided for the V-J or V-D-J junctional regions or parts thereof for the alpha and beta chains of the T-cell receptor which recognize tumor polypeptides described herein. In general, this aspect of the invention relates to T-cell receptors which recognize or bind tumor polypeptides presented in the context of MHC. In a preferred embodiment the tumor antigens recognized by the T-cell receptors comprise a polypeptide of the present invention. For example, cDNA encoding a TCR specific for a _tumor peptide can be isolated from T cells specific for a tumor polypeptide using standard molecular biological and recombinant DNA techniques.

This invention further includes the T-cell receptors or analogs thereof having substantially the same function or activity as the T-cell receptors of this invention which recognize or bind tumor polypeptides. Such receptors include, but are not limited to, a fragment of the receptor, or a substitution, addition or deletion mutant of a T-cell receptor provided herein. This invention also encompasses polypeptides or peptides that are substantially homologous to the T-cell receptors provided herein or that retain substantially the same activity. The term "analog" includes any protein or polypeptide having an amino acid residue sequence substantially identical to the T-cell receptors provided herein in which one or more residues, preferably no more than 5 residues, more

preferably no more than 25 residues have been conservatively substituted with a functionally similar residue and which displays the functional aspects of the T-cell receptor as described herein.

The present invention further provides for suitable mammalian host cells, for example, non-specific T cells, that are transfected with a polynucleotide encoding TCRs specific for a polypeptide described herein, thereby rendering the host cell specific for the polypeptide. The α and β chains of the TCR may be contained on separate expression vectors or alternatively, on a single expression vector that also contains an internal ribosome entry site (IRES) for cap-independent translation of the gene downstream of the IRES. Said host cells expressing TCRs specific for the polypeptide may be used, for example, for adoptive immunotherapy of lung cancer as discussed further below.

In further aspects of the present invention, cloned TCRs specific for a polypeptide recited herein may be used in a kit for the diagnosis of lung cancer. For example, the nucleic acid sequence or portions thereof, of tumor-specific TCRs can be used as probes or primers for the detection of expression of the rearranged genes encoding the specific TCR in a biological sample. Therefore, the present invention further provides for an assay for detecting messenger RNA or DNA encoding the TCR specific for a polypeptide.

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell and/or antibody compositions disclosed herein in pharmaceutically-acceptable carriers for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will be understood that, if desired, a composition as disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the

particular instance. Such compositions may be purified from host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

5 Therefore, in another aspect of the present invention, pharmaceutical compositions are provided comprising one or more of the polynucleotide, polypeptide, antibody, and/or T-cell compositions described herein in combination with a physiologically acceptable carrier. In certain preferred embodiments, the pharmaceutical compositions of the invention comprise immunogenic polynucleotide and/or polypeptide
10 compositions of the invention for use in prophylactic and therapeutic vaccine applications. Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Generally, such compositions will comprise one or more polynucleotide and/or polypeptide compositions of the present invention in combination with one or more immunostimulants.

15 It will be apparent that any of the pharmaceutical compositions described herein can contain pharmaceutically acceptable salts of the polynucleotides and polypeptides of the invention. Such salts can be prepared, for example, from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.*, salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (*e.g.*, sodium,
20 potassium, lithium, ammonium, calcium and magnesium salts).

 In another embodiment, illustrative immunogenic compositions, *e.g.*, vaccine compositions, of the present invention comprise DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the polynucleotide may be administered within any of a variety of delivery systems
25 known to those of ordinary skill in the art. Indeed, numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198, 1998, and references cited therein. Appropriate polynucleotide expression systems will, of course, contain the necessary regulatory DNA regulatory sequences for expression in a patient (such as a suitable promoter and terminating signal).

Alternatively, bacterial delivery systems may involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

Therefore, in certain embodiments, polynucleotides encoding immunogenic polypeptides described herein are introduced into suitable mammalian host cells for expression using any of a number of known viral-based systems. In one illustrative embodiment, retroviruses provide a convenient and effective platform for gene delivery systems. A selected nucleotide sequence encoding a polypeptide of the present invention can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to a subject. A number of illustrative retroviral systems have been described (*e.g.*, U.S. Pat. No. 5,219,740; Miller and Rosman (1989) *BioTechniques* 7:980-990; Miller, A. D. (1990) *Human Gene Therapy* 1:5-14; Scarpa et al. (1991) *Virology* 180:849-852; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; and Boris-Lawrie and Temin (1993) *Cur. Opin. Genet. Develop.* 3:102-109.

In addition, a number of illustrative adenovirus-based systems have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham (1986) *J. Virol.* 57:267-274; Bett et al. (1993) *J. Virol.* 67:5911-5921; Mittereder et al. (1994) *Human Gene Therapy* 5:717-729; Seth et al. (1994) *J. Virol.* 68:933-940; Barr et al. (1994) *Gene Therapy* 1:51-58; Berkner, K. L. (1988) *BioTechniques* 6:616-629; and Rich et al. (1993) *Human Gene Therapy* 4:461-476).

Various adeno-associated virus (AAV) vector systems have also been developed for polynucleotide delivery. AAV vectors can be readily constructed using techniques well known in the art. See, *e.g.*, U.S. Pat. Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 and WO 93/03769; Lebkowski et al. (1988) *Molec. Cell. Biol.* 8:3988-3996; Vincent et al. (1990) *Vaccines* 90 (Cold Spring Harbor Laboratory Press); Carter, B. J. (1992) *Current Opinion in Biotechnology* 3:533-539; Muzyczka, N. (1992) *Current Topics in Microbiol. and Immunol.* 158:97-129; Kotin, R. M.

(1994) Human Gene Therapy 5:793-801; Shelling and Smith (1994) Gene Therapy 1:165-169; and Zhou et al. (1994) J. Exp. Med. 179:1867-1875.

Additional viral vectors useful for delivering the polynucleotides encoding polypeptides of the present invention by gene transfer include those derived from the pox family of viruses, such as vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the novel molecules can be constructed as follows. The DNA encoding a polypeptide is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the polypeptide of interest into the viral genome. The resulting TK.sup.(-) recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

A vaccinia-based infection/transfection system can be conveniently used to provide for inducible, transient expression or coexpression of one or more polypeptides described herein in host cells of an organism. In this particular system, cells are first infected in vitro with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide or polynucleotides of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into polypeptide by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, *e.g.*, Elroy-Stein and Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-6747; Fuerst et al. Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the coding sequences of interest. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective

immunity when administered to non-avian species. The use of an Avipox vector is particularly desirable in human and other mammalian species since members of the Avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant Avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Any of a number of alphavirus vectors can also be used for delivery of polynucleotide compositions of the present invention, such as those vectors described in U.S. Patent Nos. 5,843,723; 6,015,686; 6,008,035 and 6,015,694. Certain vectors based on Venezuelan Equine Encephalitis (VEE) can also be used, illustrative examples of which can be found in U.S. Patent Nos. 5,505,947 and 5,643,576.

Moreover, molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al. J. Biol. Chem. (1993) 268:6866-6869 and Wagner et al. Proc. Natl. Acad. Sci. USA (1992) 89:6099-6103, can also be used for gene delivery under the invention.

Additional illustrative information on these and other known viral-based delivery systems can be found, for example, in Fisher-Hoch et al., *Proc. Natl. Acad. Sci. USA* 86:317-321, 1989; Flexner et al., *Ann. N.Y. Acad. Sci.* 569:86-103, 1989; Flexner et al., *Vaccine* 8:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques* 6:616-627, 1988; Rosenfeld et al., *Science* 252:431-434, 1991; Kolls et al., *Proc. Natl. Acad. Sci. USA* 91:215-219, 1994; Kass-Eisler et al., *Proc. Natl. Acad. Sci. USA* 90:11498-11502, 1993; Guzman et al., *Circulation* 88:2838-2848, 1993; and Guzman et al., *Cir. Res.* 73:1202-1207, 1993.

In certain embodiments, a polynucleotide may be integrated into the genome of a target cell. This integration may be in the specific location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the polynucleotide may be stably maintained in the cell as a separate, episomal segment of DNA. Such

polynucleotide segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. The manner in which the expression construct is delivered to a cell and where in the cell the polynucleotide remains is dependent on the type of expression construct employed.

5 In another embodiment of the invention, a polynucleotide is administered/delivered as "naked" DNA, for example as described in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

10 In still another embodiment, a composition of the present invention can be delivered via a particle bombardment approach, many of which have been described. In one illustrative example, gas-driven particle acceleration can be achieved with devices such as those manufactured by Powderject Pharmaceuticals PLC (Oxford, UK) and Powderject Vaccines Inc. (Madison, WI), some examples of which are described in U.S. Patent Nos.
15 5,846,796; 6,010,478; 5,865,796; 5,584,807; and EP Patent No. 0500 799. This approach offers a needle-free delivery approach wherein a dry powder formulation of microscopic particles, such as polynucleotide or polypeptide particles, are accelerated to high speed within a helium gas jet generated by a hand held device, propelling the particles into a target tissue of interest.

20 In a related embodiment, other devices and methods that may be useful for gas-driven needle-less injection of compositions of the present invention include those provided by Bioject, Inc. (Portland, OR), some examples of which are described in U.S. Patent Nos. 4,790,824; 5,064,413; 5,312,335; 5,383,851; 5,399,163; 5,520,639 and 5,993,412.

25 According to another embodiment, the pharmaceutical compositions described herein will comprise one or more immunostimulants in addition to the immunogenic polynucleotide, polypeptide, antibody, T-cell and/or APC compositions of this invention. An immunostimulant refers to essentially any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen.

One preferred type of immunostimulant comprises an adjuvant. Many adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis* derived proteins. Certain adjuvants are

5 commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically

10 derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may also be used as adjuvants.

Within certain embodiments of the invention, the adjuvant composition is preferably one that induces an immune response predominantly of the Th1 type. High

15 levels of Th1-type cytokines (*e.g.*, IFN- γ , TNF α , IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses.

20 Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

25 Certain preferred adjuvants for eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A, together with an aluminum salt. MPL[®] adjuvants are available from Corixa Corporation (Seattle, WA; *see*, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in

which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science* 273:352, 1996. Another
5 preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or *Chenopodium quinoa* saponins. Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β -
10 escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-
15 based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamellar liposome or ISCOM. The saponins may also be formulated with
20 excipients such as Carbopol^R to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL[®] adjuvant, as described in WO 94/00153, or a less reactogenic composition
25 where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly preferred adjuvant formulation employing QS21, 3D-MPL[®] adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and QS21 is disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

5 Additional illustrative adjuvants for use in the pharmaceutical compositions of the invention include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.*, SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Enhanzyn[®]) (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other
10 aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

Other preferred adjuvants include adjuvant molecules of the general formula
15 (I): $\text{HO}(\text{CH}_2\text{CH}_2\text{O})_n\text{-A-R}$,
wherein, n is 1-50, A is a bond or $-\text{C}(\text{O})-$, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein n is between 1 and 50, preferably 4-24, most preferably 9; the R component is C_{1-50} , preferably $\text{C}_4\text{-C}_{20}$ alkyl and
20 most preferably C_{12} alkyl, and A is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-stearyl ether, polyoxyethylene-8-stearyl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and
25 polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12th edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant

combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

According to another embodiment of this invention, an immunogenic composition described herein is delivered to a host via antigen presenting cells (APCs),
 5 such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be
 10 isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature* 392:245-251, 1998) and have been shown to be
 15 effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman and Levy, *Ann. Rev. Med.* 50:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses.
 20 Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel et al., *Nature Med.* 4:594-600, 1998).

25 Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF α to cultures of monocytes harvested from peripheral blood.

Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF α , CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

5 Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high
10 expression of Fc γ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 and 4-1BB).

15 APCs may generally be transfected with a polynucleotide of the invention (or portion or other variant thereof) such that the encoded polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a pharmaceutical composition comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets
20 a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology* 75:456-460, 1997. Antigen loading of dendritic cells may be
25 achieved by incubating dendritic cells or progenitor cells with the tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed

with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will typically vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, mucosal, intravenous, intracranial, intraperitoneal, subcutaneous and intramuscular administration.

Carriers for use within such pharmaceutical compositions are biocompatible, and may also be biodegradable. In certain embodiments, the formulation preferably provides a relatively constant level of active component release. In other embodiments, however, a more rapid rate of release immediately upon administration may be desired. The formulation of such compositions is well within the level of ordinary skill in the art using known techniques. Illustrative carriers useful in this regard include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other illustrative delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

In another illustrative embodiment, biodegradable microspheres (*e.g.*, polylactate polyglycolate) are employed as carriers for the compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344, 5,407,609 and 5,942,252. Modified hepatitis B core protein carrier systems, such as described in WO/99 40934, and references cited therein, will also be useful for many applications. Another illustrative carrier/delivery system employs a carrier comprising

particulate-protein complexes, such as those described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

The pharmaceutical compositions of the invention will often further comprise one or more buffers (*e.g.*, neutral buffered saline or phosphate buffered saline),
5 carbohydrates (*e.g.*, glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.*, aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of
10 the present invention may be formulated as a lyophilizate.

The pharmaceutical compositions described herein may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are typically sealed in such a way to preserve the sterility and stability of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily
15 or aqueous vehicles. Alternatively, a pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

The development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*,
20 oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation, is well known in the art, some of which are briefly discussed below for general purposes of illustration.

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be
25 formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions,

5 syrups, wafers, and the like (see, for example, Mathiowitz *et al.*, Nature 1997 Mar 27;386(6623):410-4; Hwang *et al.*, Crit Rev Ther Drug Carrier Syst 1998;15(3):243-84; U. S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent 5,792,451). Tablets, troches, pills, capsules and the like may also contain any of a variety of additional components, for example, a binder, such as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations will contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared in such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. Alternatively, the active ingredient may be incorporated into an oral solution such as one

containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the
5 tongue or otherwise dissolved in the mouth.

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even intraperitoneally. Such approaches are well known to the skilled artisan, some of which are further described, for example, in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S.
10 Patent 5,399,363. In certain embodiments, solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations generally will contain a preservative to prevent the
15 growth of microorganisms.

Illustrative pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (for example, see U. S. Patent 5,466,468). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists.
20 It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use
25 of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and/or by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged

absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

In one embodiment, for parenteral administration in an aqueous solution, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. Moreover, for human administration, preparations will of course preferably meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

In another embodiment of the invention, the compositions disclosed herein may be formulated in a neutral or salt form. Illustrative pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective.

The carriers can further comprise any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of

such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. The phrase

5 "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human.

In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal

10 aerosol sprays has been described, *e.g.*, in U. S. Patent 5,756,353 and U. S. Patent 5,804,212. Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*, J Controlled Release 1998 Mar 2;52(1-2):81-7) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871) are also well-known in the pharmaceutical arts. Likewise, illustrative transmucosal drug delivery in the form of a polytetrafluoroethylene

15 support matrix is described in U. S. Patent 5,780,045.

In certain embodiments, liposomes, nanocapsules, microparticles, lipid particles, vesicles, and the like, are used for the introduction of the compositions of the present invention into suitable host cells/organisms. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a

20 liposome, a vesicle, a nanosphere, or a nanoparticle or the like. Alternatively, compositions of the present invention can be bound, either covalently or non-covalently, to the surface of such carrier vehicles.

The formation and use of liposome and liposome-like preparations as potential drug carriers is generally known to those of skill in the art (see for example, Lasic, Trends Biotechnol 1998 Jul;16(7):307-21; Takakura, Nippon Rinsho 1998 Mar;56(3):691-5; Chandran *et al.*, Indian J Exp Biol. 1997 Aug;35(8):801-9; Margalit, Crit Rev Ther Drug Carrier Syst. 1995;12(2-3):233-61; U.S. Patent 5,567,434; U.S. Patent 5,552,157; U.S. Patent 5,565,213; U.S. Patent 5,738,868 and U.S. Patent 5,795,587, each specifically incorporated herein by reference in its entirety).

25

Liposomes have been used successfully with a number of cell types that are normally difficult to transfect by other procedures, including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, J Biol Chem. 1990 Sep 25;265(27):16337-42; Muller *et al.*, DNA Cell Biol. 1990 Apr;9(3):221-9). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, various drugs, radiotherapeutic agents, enzymes, viruses, transcription factors, allosteric effectors and the like, into a variety of cultured cell lines and animals. Furthermore, the use of liposomes does not appear to be associated with autoimmune responses or unacceptable toxicity after systemic delivery.

In certain embodiments, liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs)).

Alternatively, in other embodiments, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (see, for example, Quintanar-Guerrero *et al.*, Drug Dev Ind Pharm. 1998 Dec;24(12):1113-28). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) may be designed using polymers able to be degraded *in vivo*. Such particles can be made as described, for example, by Couvreur *et al.*, Crit Rev Ther Drug Carrier Syst. 1988;5(1):1-20; zur Muhlen *et al.*, Eur J Pharm Biopharm. 1998 Mar;45(2):149-55; Zambaux *et al.* J Controlled Release. 1998 Jan 2;50(1-3):31-40; and U. S. Patent 5,145,684.

Cancer Therapeutic Methods

Immunologic approaches to cancer therapy are based on the recognition that cancer cells can often evade the body's defenses against aberrant or foreign cells and molecules, and that these defenses might be therapeutically stimulated to regain the lost ground, *e.g.* pgs. 623-648 in Klein, Immunology (Wiley-Interscience, New York, 1982).

Numerous recent observations that various immune effectors can directly or indirectly inhibit growth of tumors has led to renewed interest in this approach to cancer therapy, *e.g.* Jager, et al., *Oncology* 2001;60(1):1-7; Renner, et al., *Ann Hematol* 2000 Dec;79(12):651-9.

5 Four-basic cell types whose function has been associated with antitumor cell immunity and the elimination of tumor cells from the body are: i) B-lymphocytes which secrete immunoglobulins into the blood plasma for identifying and labeling the nonself invader cells; ii) monocytes which secrete the complement proteins that are responsible for lysing and processing the immunoglobulin-coated target invader cells; iii) natural killer
10 lymphocytes having two mechanisms for the destruction of tumor cells, antibody-dependent cellular cytotoxicity and natural killing; and iv) T-lymphocytes possessing antigen-specific receptors and having the capacity to recognize a tumor cell carrying complementary marker molecules (Schreiber, H., 1989, in *Fundamental Immunology* (ed). W. E. Paul, pp. 923-955).

15 Cancer immunotherapy generally focuses on inducing humoral immune responses, cellular immune responses, or both. Moreover, it is well established that induction of CD4⁺ T helper cells is necessary in order to secondarily induce either antibodies or cytotoxic CD8⁺ T cells. Polypeptide antigens that are selective or ideally specific for cancer cells, particularly lung cancer cells, offer a powerful approach for
20 inducing immune responses against lung cancer, and are an important aspect of the present invention.

 Therefore, in further aspects of the present invention, the pharmaceutical compositions described herein may be used for the treatment of cancer, particularly for the immunotherapy of lung cancer. Within such methods, the pharmaceutical compositions
25 described herein are administered to a patient, typically a warm-blooded animal, preferably a human. A patient may or may not be afflicted with cancer. Accordingly, the above pharmaceutical compositions may be used to prevent the development of a cancer or to treat a patient afflicted with a cancer. Pharmaceutical compositions and vaccines may be administered either prior to or following surgical removal of primary tumors and/or

treatment such as administration of radiotherapy or conventional chemotherapeutic drugs. As discussed above, administration of the pharmaceutical compositions may be by any suitable method, including administration by intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intradermal, anal, vaginal, topical and oral routes.

5 Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune response-modifying agents (such as polypeptides and polynucleotides as provided herein).

10 Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established tumor-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate antitumor effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as CD8⁺ cytotoxic T lymphocytes and CD4⁺ T-helper tumor-infiltrating lymphocytes), killer
15 cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide provided herein. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The polypeptides provided herein may also be used to
20 generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive immunotherapy.

25 Monoclonal antibodies may be labeled with any of a variety of labels for desired selective usages in detection, diagnostic assays or therapeutic applications (as described in U.S. Patent Nos. 6,090,365; 6,015,542; 5,843,398; 5,595,721; and 4,708,930, hereby incorporated by reference in their entirety as if each was incorporated individually). In each case, the binding of the labelled monoclonal antibody to the determinant site of the antigen will signal detection or delivery of a particular therapeutic agent to the antigenic determinant on the non-normal cell. A further object of this invention is to provide the

specific monoclonal antibody suitably labelled for achieving such desired selective usages thereof.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast and/or B cells, may be pulsed with immunoreactive polypeptides or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. Studies have shown that cultured effector cells can be induced to grow *in vivo* and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (*see, for example, Cheever et al., Immunological Reviews 157:177, 1997*).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, intracavitary, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions described herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g., intracutaneous, intramuscular,*

intravenous or subcutaneous), intranasally (e.g., by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A
 5 suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (*i.e.*, untreated) level. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable
 10 of causing an immune response that leads to an improved clinical outcome (e.g., more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose ranges from about 25 µg to 5 mg per kg of host. Suitable
 15 dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical outcome (e.g., more
 20 frequent remissions, complete or partial, or longer disease-free survival) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a tumor protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

25 Cancer Detection and Diagnostic Compositions, Methods and Kits

In general, a cancer may be detected in a patient based on the presence of one or more lung tumor proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum urine and/or tumor biopsies) obtained

from the patient. In other words, such proteins may be used as markers to indicate the presence or absence of a cancer such as lung cancer. In addition, such proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample.

5 Polynucleotide primers and probes may be used to detect the level of mRNA encoding a tumor protein, which is also indicative of the presence or absence of a cancer. In general, a tumor sequence should be present at a level that is at least two-fold, preferably three-fold, and more preferably five-fold or higher in tumor tissue than in normal tissue of the same type from which the tumor arose. Expression levels of a particular tumor
10 sequence in tissue types different from that in which the tumor arose are irrelevant in certain diagnostic embodiments since the presence of tumor cells can be confirmed by observation of predetermined differential expression levels, e.g., 2-fold, 5-fold, etc, in tumor tissue to expression levels in normal tissue of the same type.

Other differential expression patterns can be utilized advantageously for
15 diagnostic purposes. For example, in one aspect of the invention, overexpression of a tumor sequence in tumor tissue and normal tissue of the same type, but not in other normal tissue types, e.g. PBMCs, can be exploited diagnostically. In this case, the presence of metastatic tumor cells, for example in a sample taken from the circulation or some other tissue site different from that in which the tumor arose, can be identified and/or confirmed
20 by detecting expression of the tumor sequence in the sample, for example using RT-PCR analysis. In many instances, it will be desired to enrich for tumor cells in the sample of interest, e.g., PBMCs, using cell capture or other like techniques.

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. See, e.g., Harlow
25 and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, the presence or absence of a cancer in a patient may be determined by (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

In a preferred embodiment, the assay involves the use of binding agent immobilized on a solid support to bind to and remove the polypeptide from the remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex.

5 Such detection reagents may comprise, for example, a binding agent that specifically binds to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G, protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent
10 with the sample. The extent to which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding agent. Suitable polypeptides for use within such assays include full length lung tumor proteins and polypeptide portions thereof to which the binding agent binds, as described above.

15 The solid support may be any material known to those of ordinary skill in the art to which the tumor protein may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic
20 particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681. The binding agent may be immobilized on the solid support using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which
25 may be a direct linkage between the agent and functional groups on the support or may be a linkage by way of a cross-linking agent). Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1

hour and about 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10 μ g, and preferably about 100 ng to about 1 μ g, is sufficient to immobilize an adequate amount of binding agent.

5 Covalent attachment of binding agent to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group
10 on the support with an amine and an active hydrogen on the binding partner (*see, e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that
15 polypeptides within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the
20 specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO). The immobilized
25 antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with lung cancer. Preferably, the contact time is

sufficient to achieve a level of binding that is at least about 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide. An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of a cancer, such as lung cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average mean signal obtained when the immobilized antibody is incubated with samples from patients without the cancer. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve,

according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the binding agent is immobilized on a membrane, such as nitrocellulose. In the flow-through test, polypeptides within the sample bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution containing the second binding agent flows through the membrane. The detection of bound second binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. Concentration of second binding agent at the area of immobilized antibody indicates the presence of a cancer. Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody sandwich assay, in the format discussed above. Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of

antibody immobilized on the membrane ranges from about 25 ng to about 1 μ g, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use tumor polypeptides to detect antibodies that bind to such polypeptides in a biological sample. The detection of such tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a tumor protein in a biological sample. Within certain methods, a biological sample comprising CD4⁺ and/or CD8⁺ T cells isolated from a patient is incubated with a tumor polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with polypeptide (*e.g.*, 5 - 25 μ g/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of tumor polypeptide to serve as a control. For CD4⁺ T cells, activation is preferably detected by evaluating proliferation of the T cells. For CD8⁺ T cells, activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a tumor protein in a biological sample. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a tumor cDNA derived from a biological sample, wherein at

least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the tumor protein. The amplified cDNA is then separated and detected using techniques well known in the art, such as gel electrophoresis.

Similarly, oligonucleotide probes that specifically hybridize to a
 5 polynucleotide encoding a tumor protein may be used in a hybridization assay to detect the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a
 10 polynucleotide encoding a tumor protein of the invention that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are
 15 at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence as disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.*, 51:263, 1987; Erlich ed.,
 20 *PCR Technology*, Stockton Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be
 25 separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in

expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another aspect of the present invention, cell capture technologies may be used in conjunction, with, for example, real-time PCR to provide a more sensitive tool for
 5 detection of metastatic cells expressing lung tumor antigens. Detection of lung cancer cells in biological samples, e.g., bone marrow samples, peripheral blood, and small needle aspiration samples is desirable for diagnosis and prognosis in lung cancer patients.

Immunomagnetic beads coated with specific monoclonal antibodies to surface cell markers, or tetrameric antibody complexes, may be used to first enrich or
 10 positively select cancer cells in a sample. Various commercially available kits may be used, including Dynabeads® Epithelial Enrich (DynaL Biotech, Oslo, Norway), StemSep™ (StemCell Technologies, Inc., Vancouver, BC), and RosetteSep (StemCell Technologies). A skilled artisan will recognize that other methodologies and kits may also be used to enrich or positively select desired cell populations. Dynabeads® Epithelial Enrich contains
 15 magnetic beads coated with MAbs specific for two glycoprotein membrane antigens expressed on normal and neoplastic epithelial tissues. The coated beads may be added to a sample and the sample then applied to a magnet, thereby capturing the cells bound to the beads. The unwanted cells are washed away and the magnetically isolated cells eluted from the beads and used in further analyses.

20 RosetteSep can be used to enrich cells directly from a blood sample and consists of a cocktail of tetrameric antibodies that targets a variety of unwanted cells and crosslinks them to glycophorin A on red blood cells (RBC) present in the sample, forming rosettes. When centrifuged over Ficoll, targeted cells pellet along with the free RBC. The combination of antibodies in the depletion cocktail determines which cells will be removed
 25 and consequently which cells will be recovered. Antibodies that are available include, but are not limited to: CD2, CD3, CD4, CD5, CD8, CD10, CD11b, CD14, CD15, CD16, CD19, CD20, CD24, CD25, CD29, CD33, CD34, CD36, CD38, CD41, CD45, CD45RA, CD45RO, CD56, CD66B, CD66e, HLA-DR, IgE, and TCR $\alpha\beta$.

Additionally, it is contemplated in the present invention that MAbs specific for lung tumor antigens can be generated and used in a similar manner. For example, MAbs that bind to tumor-specific cell surface antigens may be conjugated to magnetic beads, or formulated in a tetrameric antibody complex, and used to enrich or positively
 5 select metastatic lung tumor cells from a sample. Once a sample is enriched or positively selected, cells may be lysed and RNA isolated. RNA may then be subjected to RT-PCR analysis using lung tumor-specific primers in a real-time PCR assay as described herein. One skilled in the art will recognize that enriched or selected populations of cells may be analyzed by other methods (*e.g. in situ* hybridization or flow cytometry).

10 In another embodiment, the compositions described herein may be used as markers for the progression of cancer. In this embodiment, assays as described above for the diagnosis of a cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter performed
 15 as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor.
 20 One such assay involves contacting tumor cells with a binding agent. The bound binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple tumor protein markers may
 25 be assayed within a given sample. It will be apparent that binding agents specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in

optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a tumor protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a tumor protein. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1

IDENTIFICATION AND CHARACTERIZATION OF LUNG TUMOR cDNAS

This Example illustrates the identification of cDNA molecules encoding lung tumor proteins.

A. Isolation of cDNA Sequences from Lung Adenocarcinoma Libraries using Conventional cDNA Library Subtraction

A human lung adenocarcinoma cDNA expression library was constructed from poly A⁺ RNA from patient tissues (# 40031486) using a Superscript Plasmid System for cDNA Synthesis and Plasmid Cloning kit (BRL Life Technologies, Gaithersburg, MD) following the manufacturer's protocol. Specifically, lung carcinoma tissues were homogenized with polytron (Kinematica, Switzerland) and total RNA was extracted using Trizol reagent (BRL Life Technologies) as directed by the manufacturer. The poly A⁺ RNA was then purified using an oligo dT cellulose column as described in Sambrook et al.,
 10 *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. First-strand cDNA was synthesized using the NotI/Oligo-dT18 primer. Double-stranded cDNA was synthesized, ligated with BstXI/EcoRI adaptors (Invitrogen, San Diego, CA) and digested with NotI. Following size fractionation with cDNA size fractionation columns (BRL Life Technologies), the cDNA was ligated into the BstXI/NotI
 15 site of pcDNA3.1 (Invitrogen) and transformed into ElectroMax *E. coli* DH10B cells (BRL Life Technologies) by electroporation. A total of 3×10^6 independent colonies were generated.

Using the same procedure, a normal human cDNA expression library was prepared from a panel of normal tissue specimens, including lung, liver, pancreas, skin,
 20 kidney, brain and resting PBMC.

cDNA library subtraction was performed using the above lung adenocarcinoma and normal tissue cDNA libraries, as described by Hara *et al.* (*Blood*, 84:189-199, 1994) with some modifications. Specifically, a lung adenocarcinoma-specific subtracted cDNA library was generated as follows. The normal tissue cDNA library
 25 (80 µg) was digested with BamHI and XhoI, followed by a filling-in reaction with DNA polymerase Klenow fragment. After phenol-chloroform extraction and ethanol precipitation, the DNA was dissolved in 133 µl of H₂O, heat-denatured and mixed with 133 µl (133 µg) of Photoprobe biotin (Vector Laboratories, Burlingame, CA). As recommended by the manufacturer, the resulting mixture was irradiated with a 270 W

sunlamp on ice for 20 minutes. Additional Photoprobe biotin (67 μ l) was added and the biotinylation reaction was repeated. After extraction with butanol five times, the DNA was ethanol-precipitated and dissolved in 23 μ l H₂O. The resulting DNA, plus other highly redundant cDNA clones that were frequently recovered in previous lung subtractions

5 formed the driver DNA.

To form the tracer DNA, 10 μ g lung adenocarcinoma cDNA library was digested with NotI and SpeI, phenol chloroform extracted and passed through Chroma spin-400 columns (Clontech, Palo Alto, CA). Typically, 5 μ g of cDNA was recovered after the sizing column. Following ethanol precipitation, the tracer DNA was dissolved in 5 μ l H₂O.

10 Tracer DNA was mixed with 15 μ l driver DNA and 20 μ l of 2 x hybridization buffer (1.5 M NaCl/10 mM EDTA/50 mM HEPES pH 7.5/0.2% sodium dodecyl sulfate), overlaid with mineral oil, and heat-denatured completely. The sample was immediately transferred into a 68 °C water bath and incubated for 20 hours (long hybridization [LH]). The reaction mixture was then subjected to a streptavidin treatment followed by phenol/chloroform

15 extraction. This process was repeated three more times. Subtracted DNA was precipitated, dissolved in 12 μ l H₂O, mixed with 8 μ l driver DNA and 20 μ l of 2 x hybridization buffer, and subjected to a hybridization at 68°C for 2 hours (short hybridization [SH]). After removal of biotinylated double-stranded DNA, subtracted cDNA was ligated into NotI/SpeI site of chloramphenicol resistant pBCSK⁺ (Stratagene, La Jolla, CA) and transformed into

20 ElectroMax *E. coli* DH10B cells by electroporation to generate a lung adenocarcinoma specific subtracted cDNA library, referred to as LAT-S1. Similarly, LAT-S2 was generated by including 23 genes that were over-expressed in the tracer as additional drivers.

A second human lung adenocarcinoma cDNA expression library was constructed using adenocarcinoma tissue from a second patient (# 86-66) and used to

25 prepare a second lung adenocarcinoma-specific subtracted cDNA library (referred to as LAT2-S2), as described above, using the same panel of normal tissues and the additional genes over-expressed in LAT-S1.

A third human metastatic lung adenocarcinoma library was constructed from a pool of two lung pleural effusions with lung and gastric adenocarcinoma origins. The

subtracted cDNA library, referred to as Mets-sub2, was generated as described above using the same panel of normal tissues. The Mets-sub3 subtracted library was constructed by including 51 additional genes as drivers. These 51 genes were recovered in Mets-sub2, representing over-expressed housekeeping genes in the testers.

- 5 A total of 16 cDNA fragments isolated from LAT-S1, 585 cDNA fragments isolated from LAT-S2, 568 cDNA clones from LAT2-S2, 15 cDNA clones from Mets-sub2 and 343 cDNA clones from Mets-sub3, described above, were colony PCR amplified and their mRNA expression levels in lung tumor, normal lung, and various other normal and tumor tissues were determined using microarray technology (Incyte, Palo Alto, CA).
- 10 Briefly, the PCR amplification products were dotted onto slides in an array format, with each product occupying a unique location in the array. mRNA was extracted from the tissue sample to be tested, reverse transcribed, and fluorescent-labeled cDNA probes were generated. The microarrays were probed with the labeled cDNA probes, the slides scanned and fluorescence intensity was measured. This intensity correlates with the hybridization
- 15 intensity. Seventy-three non-redundant cDNA clones, of which 42 were found to be unique, showed over-expression in lung tumors, with expression in normal tissues tested (lung, skin, lymph node, colon, liver, pancreas, breast, heart, bone marrow, large intestine, kidney, stomach, brain, small intestine, bladder and salivary gland) being either undetectable, or at significantly lower levels compared to lung adenocarcinoma tumors.
- 20 These clones were further characterized by DNA sequencing with a Perkin Elmer/Applied Biosystems Division Automated Sequencer Model 373A and/or Model 377 (Foster City, CA).

- 25 The sequences were compared to known sequences in the gene bank using the EMBL GenBank databases (release 96). No significant homologies were found to the sequence provided in SEQ ID NO:67, with no apparent homology to previously identified expressed sequence tags (ESTs). The sequences of SEQ ID NO:60, 62, 65, 66, 69-71, 74, 76, 79, 80, 84, 86, 89-92, 95, 97 and 98 were found to show some homology to previously identified expressed sequence tags (ESTs). The cDNA sequences of SEQ ID NO:59, 61, 63, 64, 67, 68, 72, 73, 75, 77, 78, 81-83, 85, 87, 88, 93, 94, 96, 99 and 100 showed

homology to previously identified genes. The full-length cDNA sequences for the clones of SEQ ID NO:96 and 100 are provided in SEQ ID NO:316 and 318, respectively. The amino acid sequences for the clones of SEQ ID NO: 59, 61, 63, 64, 68, 73, 82, 83, 94, 96 and 100 are provided in SEQ ID NO:331, 328, 329, 332, 327, 333, 330, 326, 325, 324 and
 5 335, respectively. The amino acid sequence encoded by the sequence of SEQ ID NO:69 (referred to as L552S) is provided in SEQ ID NO:786.

Further studies led to the isolation of an extended cDNA sequence and the open reading frame for L552S (SEQ ID NO:790). The amino acid sequence encoded by the cDNA sequence of SEQ ID NO:790 is provided in SEQ ID NO:791. Subsequent studies
 10 led to the isolation of the full-length cDNA sequence of L552S (SEQ ID NO:808). The full-length cDNA of L552S has an open-reading frame of 480 base pairs (SEQ ID NO:790) and encodes a putative polypeptide of 160 amino acids (SEQ ID NO:809).

Initial database searches failed to detect any sequence homology with proteins in the database, suggesting that L552S encodes a novel protein of unknown
 15 function. Recently, a cancer-testis antigen, XAGE-1, was found to be over-expressed in Ewing's Sarcoma (Liu *et al.*, 2000 Cancer Res. 60:4752-4755). The determined cDNA sequence of XAGE-1 is provided in SEQ ID NO:792 with the corresponding amino acid sequence being provided in SEQ ID NO:793. A sequence comparison of L552S and XAGE-1 reveals striking identities as well as differences. The majority of the C-terminal
 20 sequences are identical to each other. The polypeptides predicted from L552S and XAGE-1 have diverged N-terminal sequences. Hydrophilicity analysis of the L552S amino acids suggested a very hydrophilic protein with no transmembrane domains predicted. PSORT analysis of L552S revealed the same result. Since L552S is also localized in chromosome X, it is likely to be a new isoform of cancer testis antigen, XAGE-1. The genomic
 25 sequence analysis revealed that both genes localized in the same region of the X chromosome and both have four exons. The last three exons are identical for L552S and XAGE-1. However, the first exon for XAGE-1 is upstream of the first exon for L552S and this results in distinct 5' nucleotide and amino acid sequences for L552S and XAGE-1. Therefore, L552S and XAGE-1 are alternatively spliced isoforms.

The amino acid sequence of this N-terminal portion of L552S is provided in SEQ ID NO:1830 with the corresponding cDNA sequence being provided in SEQ ID NO:1826. The cDNA sequences provided in SEQ ID NO:1827-1829 represent bp 394-681 of SEQ ID NO:808, bp394-534 of SEQ ID NO:808 and bp 214-394 of SEQ ID NO:792, 5 respectively, with the corresponding amino acid sequences being provided in SEQ ID NO:1831-1833, respectively.

Full-length cloning efforts on L552S also led to the isolation of three additional cDNA sequences (SEQ ID NO:810-812; referred to as clones 50989, 50990 and 50992, respectively) from a metastatic lung adenocarcinoma library. The sequence of SEQ 10 ID NO:810 was found to show some homology to previously identified human DNA sequences. The sequence of SEQ ID NO:811 was found to show some homology to a previously identified DNA sequence. The sequence of SEQ ID NO:812 was found to show some homology to previously identified ESTs.

The gene of SEQ ID NO:84 (referred to as L551S) was determined by real- 15 time RT-PCR analysis to be over-expressed in 2/9 primary adenocarcinomas and to be expressed at lower levels in 2/2 metastatic adenocarcinomas and 1/2 squamous cell carcinomas. No expression was observed in normal tissues, with the exception of very low expression in normal stomach. Further studies on L551S led to the isolation of the 5' and 3' cDNA consensus sequences provided in SEQ ID NO:801 and 802, respectively. The 20 L551S 5' sequence was found to show some homology to the previously identified gene STY8 (cDNA sequence provided in SEQ ID NO:803; corresponding amino acid sequence provided in SEQ ID NO:805), which is a mitogen activated protein kinase phosphatase. However, no significant homologies were found to the 3' sequence of L551S. Subsequently, an extended cDNA sequence for L551S was isolated (SEQ ID NO:804). 25 The corresponding amino acid sequence is provided in SEQ ID NO:806. Further studies led to the isolation of two independent full-length clones for L551S (referred to as 54298 and 54305). These two clones have five nucleotide differences compared to the STY8 DNA sequence. Two of these differences are single nucleotide polymorphisms which do not effect the encoded amino acid sequences. The other three nucleotide differences are

consistent between the two L551S clones but lead to encoded amino acid sequences that are different from the STY8 protein sequence. The determined cDNA sequences for the L551S full-length clones 54305 and 54298 are provided in SEQ ID NO:825 and 826, respectively, with the amino acid sequence for L551S being provided in SEQ ID NO:827.

5 B. Isolation of cDNA Sequences from Lung Adenocarcinoma Libraries using PCR-Based cDNA Library Subtraction

cDNA clones from a subtracted library, containing cDNA from a pool of two human lung primary adenocarcinomas subtracted against a pool of nine normal human tissue cDNAs including skin, colon, lung, esophagus, brain, kidney, spleen, pancreas and
 10 liver, (Clontech, Palo Alto, CA) were submitted to a first round of PCR amplification. This library (referred to as ALT-1) was subjected to a second round of PCR amplification, following the manufacturer's protocol. The expression levels of 760 cDNA clones in lung tumor, normal lung, and various other normal and tumor tissues, were examined using microarray technology as described above. A total of 118 clones, of which 55 were unique,
 15 were found to be over-expressed in lung tumor tissue, with expression in normal tissues tested (lung, skin, lymph node, colon, liver, pancreas, breast, heart, bone marrow, large intestine, kidney, stomach, brain, small intestine, bladder and salivary gland) being either undetectable, or at significantly lower levels. The sequences were compared to known sequences in the gene bank using the EMBL and GenBank databases (release 96). No
 20 significant homologies (including ESTs) were found to the sequence provided in SEQ ID NO:44. The sequences of SEQ ID NO:1, 11, 13, 15, 20, 23-27, 29, 30, 33, 34, 39, 41, 43, 45, 46, 51 and 57 were found to show some homology to previously identified expressed sequence tags (ESTs). The cDNA sequences of SEQ ID NO:2-10, 12, 14, 16-19, 21, 22, 28, 31, 32, 35-38, 40, 42, 44, 47-50, 52-56 and 58 showed homology to previously identified
 25 genes. The full-length cDNA sequences for the clones of SEQ ID NO:18, 22, 31, 35, 36 and 42 are provided in SEQ ID NO:320, 319, 323, 321, 317, 321 and 322, respectively, with the corresponding amino acid sequences being provided in SEQ ID NO: 337, 336, 340, 338, 334, and 339, respectively.

Further studies led to the isolation of an extended cDNA sequence for the clone of SEQ ID NO:33 (referred to as L801P). This extended cDNA sequence (provided in SEQ ID NO:796), was found to contain three potential open reading frames (ORFs). The predicted amino acid sequences encoded by these three ORFs are provided in SEQ ID
 5 NO:797-799, respectively. Additional full-length cloning efforts led to still further extended cDNA sequence for L801P, set forth in SEQ ID NO:1669, in addition to five potential open reading frames (referred to as ORFs 4-9; SEQ ID NO:1670-1675, respectively) encoded by the extended cDNA sequence. L801P was mapped to chromosomal region 20p13 and a 137 amino acid ORF from this genomic region was
 10 identified that corresponds to ORF4 (SEQ ID NO:1670), suggesting that this is likely an authentic ORF for L801P.

By microarray analysis, L801P was found to be overexpressed by 2-fold or greater in lung tumor tissue compared to normal tissue. By real-time PCR analysis, greater than 50% of lung adenocarcinoma and greater than 30% of lung squamous cell carcinoma
 15 tumor samples tested had elevated L801P expression relative to normal lung tissue. Of those that displayed elevated L801P, the level of expression was greater than 10-fold higher than in normal lung tissue samples. Moreover, low or no expression of L801P was detected in an extensive panel of normal tissue RNAs.

L801P expression was also detected in a number of other tumor types,
 20 including breast, prostate, ovarian and colon tumors, and thus may have diagnostic and/or therapeutic utility in these cancer types.

In subsequent studies, a full-length cDNA sequence for the clone of SEQ ID NO:44 (referred to as L844P) was isolated (provided in SEQ ID NO:800). Comparison of this sequence with those in the public databases revealed that the 470 bases at the 5' end of
 25 the sequence show homology to the known gene dihydrodiol dehydrogenase, thus indicating that L844P is a novel transcript of the dihydrodiol dehydrogenase family having 2007 base pairs of previously unidentified 3' untranslated region.

The predicted amino acid sequence encoded by the sequence of SEQ ID NO:46 (referred to as L840P) is provided in SEQ ID NO:787. An extended cDNA

sequence for L840P, which was determined to include an open reading frame, is provided in SEQ ID NO:794. The predicted amino acid sequence encoded by the cDNA sequence of SEQ ID NO:794 is provided in SEQ ID NO:795. The full-length cDNA sequence for the clone of SEQ ID NO:54 (referred to as L548S) is provided in SEQ ID NO:788, with the
 5 corresponding amino acid sequence being provided in SEQ ID NO:789.

Northern blot analyses of the genes of SEQ ID NO:25 and 46 (referred to as L839P and L840P, respectively) were remarkably similar. Both genes were expressed in 1/2 lung adenocarcinomas as two bands of 3.6 kb and 1.6 kb. No expression of L839P was observed in normal lung or trachea. No expression of L840P was observed in normal bone
 10 marrow, resting or activated PBMC, esophagus, or normal lung. Given the similar expression patterns, L839P and L840P may be derived from the same gene.

Additional lung adenocarcinoma cDNA clones were isolated as follows. A cDNA library was prepared from a pool of two lung adenocarcinomas and subtracted against cDNA from a panel of normal tissues including lung, brain, liver, kidney, pancreas,
 15 skin, heart and spleen. The subtraction was performed using a PCR-based protocol (Clontech), which was modified to generate larger fragments. Within this protocol, tester and driver double stranded cDNA were separately digested with five restriction enzymes that recognize six-nucleotide restriction sites (MluI, MscI, PvuII, SalI and StuI). This digestion resulted in an average cDNA size of 600 bp, rather than the average size of 300
 20 bp that results from digestion with RsaI according to the Clontech protocol. The ends of the restriction digested tester cDNA were filled in to generate blunt ends for adapter ligation. This modification did not affect the subtraction efficiency. Two tester populations were then created with different adapters, and the driver library remained without adapters. The tester and driver libraries were then hybridized using excess driver
 25 cDNA. In the first hybridization step, driver was separately hybridized with each of the two tester cDNA populations. This resulted in populations of (a) unhybridized tester cDNAs, (b) tester cDNAs hybridized to other tester cDNAs, (c) tester cDNAs hybridized to driver cDNAs and (d) unhybridized driver cDNAs. The two separate hybridization reactions were then combined, and rehybridized in the presence of additional denatured

driver cDNA. Following this second hybridization, in addition to populations (a) through (d), a fifth population (e) was generated in which tester cDNA with one adaptor hybridized to tester cDNA with the second adaptor. Accordingly, the second hybridization step resulted in enrichment of differentially expressed sequences which could be used as
5 templates for PCR amplification with adaptor-specific primers.

The ends were then filled in, and PCR amplification was performed using adaptor-specific primers. Only population (e), which contained tester cDNA that did not hybridize to driver cDNA, was amplified exponentially. A second PCR amplification step was then performed, to reduce background and further enrich differentially expressed
10 sequences.

Fifty-seven cDNA clones were isolated from the subtracted library (referred to as LAP1) and sequenced. The determined cDNA sequences for 16 of these clones are provided in SEQ ID NO:101-116. The sequences of SEQ ID NO:101 and 114 showed no significant homologies to previously identified sequences. The sequences of SEQ ID
15 NO:102-109 and 112 showed some similarity to previously identified sequences, while the sequences of SEQ ID NO:113, 115 and 116 showed some similarity to previously isolated ESTs.

An additional 502 clones analyzed from the LAP1 library were sequenced and the determined cDNA sequences are shown in SEQ ID NO:828-1239 and 1564-1653.

20 C. Isolation of cDNA Sequences from Small Cell Lung Carcinoma Libraries using PCR-Based cDNA Library Subtraction

A subtracted cDNA library for small cell lung carcinoma (referred to as SCL1) was prepared essentially using the modified PCR-based subtraction process described above. cDNA from small cell lung carcinoma was subtracted against cDNA
25 from a panel of normal tissues, including normal lung, brain, kidney, liver, pancreas, skin, heart, lymph node and spleen. Both tester and driver poly A+ RNA were initially amplified using SMART PCR cDNA synthesis kit (Clontech, Palo Alto, CA). The tester and driver double stranded cDNA were separately digested with five restriction enzymes (DraI, MscI,

PvuII, SmaI, and StuI). These restriction enzymes generated blunt end cuts and the digestion resulted in an average insert size of 600 bp. Digestion with this set of restriction enzymes eliminates the step required to generate blunt ends by filling in of the cDNA ends. These modifications did not affect subtraction efficiency.

5 Eighty-five clones were isolated and sequenced. The determined cDNA sequences for 31 of these clones are provided in SEQ ID NO:117-147. The sequences of SEQ ID NO:122, 124, 126, 127, 130, 131, 133, 136, 139 and 147 showed no significant homologies to previously identified sequences. The sequences of SEQ ID NO:120, 129, 135, 137, 140, 142, 144 and 145 showed some similarity to previously identified gene
10 sequences, while the sequences of SEQ ID NO:114, 118, 119, 121, 123, 125, 128, 132, 134, 138, 141, 143 and 147 showed some similarity to previously isolated ESTs.

In further studies, three additional cDNA libraries were generated from poly A+ RNA from a single small cell lung carcinoma sample subtracted against a pool of poly A+ RNA from nine normal tissues (lung, brain, kidney, liver, pancreas, skin, heart pituitary
15 gland and spleen). For the first library (referred to as SCL2), the subtraction was carried out essentially as described above for the LAP1 library, with the exception that the tester and driver were digested with PvuII, StuI, MscI and DraI. The ratio of tester and driver cDNA used was as recommended by Clontech. For the second library (referred to as SCL3), subtraction was performed essentially as for SCL2 except that cDNA for highly
20 redundant clones identified from the SCL2 library was included in the driver cDNA. Construction of the SCL4 library was performed essentially as described for the SCL3 library except that a higher ratio of driver to tester was employed.

Each library was characterized by DNA sequencing and database analyses. The determined cDNA sequence for 35 clones isolated from the SCL2 library are provided
25 in SEQ ID NO:245-279, with the determined cDNA sequences for 21 clones isolated from the SCL3 library and for 15 clones isolated from the SCL4 library being provided in SEQ ID NO:280-300 and 301-315, respectively. The sequences of SEQ ID NO:246, 254, 261, 262, 304, 309 and 311 showed no significant homologies to previously identified sequences. The sequence of SEQ ID NO:245, 248, 255, 266, 270, 275, 280, 282, 283, 288-

290, 292, 295, 301 and 303 showed some homology to previously isolated ESTs, while the sequences of SEQ ID NO:247, 249-253, 256-260, 263-265, 267-269, 271-274, 276-279, 281, 284-287, 291, 293, 294, 296-300, 302, 305-308, 310 and 312-315 showed some homology to previously identified gene sequences.

- 5 Sequences disclosed herein were further evaluated for overexpression in specific tumor tissues by microarray analysis. Using this approach, cDNA sequences were PCR amplified and their mRNA expression profiles in tumor and normal tissues were examined using cDNA microarray technology essentially as described (Shena, M. *et al.*, 1995 Science 270:467-70). In brief, the clones were arrayed onto glass slides as multiple
- 10 replicas, with each location corresponding to a unique cDNA clone (as many as 5500 clones can be arrayed on a single slide or chip). Each chip was hybridized with a pair of cDNA probes that are fluorescence-labeled with Cy3 and Cy5, respectively. Typically, 1µg of polyA⁺ RNA was used to generate each cDNA probe. After hybridization, the chips were scanned and the fluorescence intensity recorded for both Cy3 and Cy5 channels.
- 15 There were multiple built-in quality control steps. First, the probe quality was monitored using a panel of ubiquitously expressed genes. Secondly, the control plate also can include yeast DNA fragments of which complementary RNA may be spiked into the probe synthesis for measuring the quality of the probe and the sensitivity of the analysis. Currently, the technology offers a sensitivity of 1 in 100,000 copies of mRNA. Finally, the
- 20 reproducibility of this technology can be ensured by including duplicated control cDNA elements at different locations.

- 3264 cDNA clones from three PCR-based subtracted cDNA libraries were analyzed by the above cDNA microarray technology. The cDNA clones were arrayed on Lung Chip 5. Of these cDNA clones, 960 clones came from SQL1 library, 768 clones
- 25 came from SCL1 library, and 1536 clones came from SCL3 and SCL4 libraries. Thirty-five pairs of fluorescent labeled cDNA probes were used for the microarray analysis. Each probe pair included a lung tumor probe paired with a normal tissue probe. The expression data was analyzed. 498 cDNA clones were found to be overexpressed by 2-fold or greater in the small cell and/or non-small cell lung tumor probe groups compared to the normal

tissue probe group. Also, the mean expression values for these clones in normal tissues were below 0.1 (range of expression is from 0.001 to 10). The cDNA sequences disclosed in SEQ ID NO:1240-1563 represent 324 non-redundant clones.

The following sequences were novel based on database analysis including
 5 GenBank and GeneSeq: SEQ ID NOs:1240, 1243, 1247, 1269, 1272, 1280, 1283, 1285, 1286, 1289, 1300, 1309, 1318, 1319, 1327, 1335, 1339, 1346, 1359, 1369, 1370, 1371, 1393, 1398, 1405, 1408, 1413, 1414, 1417, 1422, 1429, 1432, 1435, 1436, 1438-1442, 1447, 1450, 1453, 1463, 1467, 1470, 1473, 1475, 1482, 1486, 1491-1494, 1501, 1505, 1506, 1514-1517, 1520, 1522, 1524, 1535, 1538, 1542, 1543, 1547, 1554, 1557, 1559,
 10 1561, and 1563.

The extended cDNA sequence of the partial sequence of contig 139 (SEQ ID NO:1467), also known as L985P, was predicted by searching public databases using SEQ ID NO:1467 as a query. By this approach, it was found that SEQ ID NO:1467 had
 15 homology to a cDNA sequence (SEQ ID NO:1676) which encodes the cell surface immunomodulator-2 (CSIMM-2). The cDNA sequence of SEQ ID NO:1676 encodes a protein having the sequence set forth in SEQ ID NO:1677.

By microarray analysis, L985P was overexpressed by 2-fold or greater in the lung tumor probe groups compared to the normal tissue probe group. Moreover, the mean expression values for L985P in normal tissues was below 0.2 (range of expression was
 20 from 0.01 to 10). By real-time PCR analysis, greater than 40% of small cell lung carcinoma lung tumor samples tested had elevated L985P expression relative to normal lung tissue. Of those that displayed elevated L985P, the level of expression was greater than 3-fold higher than in normal lung tissue samples. Low or no expression of L985P was detected in an extensive panel of normal tissue RNAs. These findings for L985P support
 25 its use both as a diagnostic marker for detecting the presence of lung cancer in a patient and/or as an immunotherapeutic target for the treatment of lung cancer.

D. Isolation of cDNA Sequences from a Neuroendocrine Library using PCR-Based cDNA Library Subtraction

Using the modified PCR-based subtraction process, essentially as described above for the LAP1 subtracted library, a subtracted cDNA library (referred to as MLN1) was derived from a lung neuroendocrine carcinoma that had metastasized to the subcarinal lymph node, by subtraction with a panel of nine normal tissues, including normal lung, brain, kidney, liver, pancreas, skin, heart, lymph node and spleen.

Ninety-one individual clones were isolated and sequenced. The determined cDNA sequences for 58 of these clones are provided in SEQ ID NO:147-222. The sequences of SEQ ID NO:150, 151, 154, 157, 158, 159, 160, 163, 174, 175, 178, 186-190, 192, 193, 195-200, 208-210, 212-215 and 220 showed no significant homologies to previously identified sequences. The sequences of SEQ ID NO:152, 155, 156, 161, 165, 166, 176, 179, 182, 184, 185, 191, 194, 221 and 222 showed some similarity to previously identified gene sequences, while the sequences of SEQ ID NO:148, 149, 153, 164, 167-173, 177, 180, 181, 183, 201-207, 211 and 216-219 showed some similarity to previously isolated ESTs.

The determined cDNA sequences of an additional 442 clones isolated from the MLN1 library are provided in SEQ ID NO:341-782, with the determined cDNA sequences of an additional 11 clones isolated from the MLN1 library are provided in SEQ ID NO:1654-1664.

E. Isolation of cDNA Sequences from a Squamous Cell Lung Carcinoma Library using PCR-Based cDNA Library Subtraction

A subtracted cDNA library for squamous cell lung carcinoma (referred to as SQL1) was prepared essentially using the modified PCR-based subtraction process described above, except the tester and driver double stranded cDNA were separately digested with four restriction enzymes (DraI, MscI, PvuII and StuI). cDNA from a pool of two squamous cell lung carcinomas was subtracted against cDNA from a pool of 10 normal

tissues, including normal lung, brain, kidney, liver, pancreas, skin, heart, spleen, esophagus and trachea.

Seventy-four clones were isolated and sequenced. The determined cDNA sequences for 22 of these clones are provided in SEQ ID NO:223-244. The sequence of
 5 SEQ ID NO:241 showed no significant homologies to previously identified sequences. The sequences of SEQ ID NO:223, 225, 232, 233, 235, 238, 239, 242 and 243 showed some similarity to previously identified gene sequences, while the sequences of SEQ ID NO:224, 226-231, 234, 236, 237, 240, 241 and 244 showed some similarity to previously isolated ESTs.

10 The sequences of an additional 12 clones isolated during characterization of cDNA libraries prepared from lung tumor tissue are provided in SEQ ID NO:813-824. Comparison of these sequences with those in the GenBank database and the GeneSeq DNA database revealed no significant homologies to previously identified sequences.

15

EXAMPLE 2

SYNTHESIS OF POLYPEPTIDES

Polypeptides may be synthesized on a Perkin Elmer/Applied Biosystems Division 430A peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-
 20 N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation, binding to an immobilized surface, or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water :phenol (40:1:2:2:3). After cleaving for
 25 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides.

Following lyophilization of the pure fractions, the peptides may be characterized using electrospray or other types of mass spectrometry and by amino acid analysis.

5

EXAMPLE 3

EXPRESSION IN *E. COLI* OF L548S HIS TAG FUSION PROTEIN

The L548S coding region was PCR amplified with the following primers:

10 Forward primer starting at amino acid 2:

PDM-433: 5' gctaaaggtgaccccaagaaaccaaag 3' Tm 60°C (SEQ ID NO:1665)

Reverse primer creating a XhoI site after the stop codon:

PDM-438: 5' ctattaactcgagggagacagataaacagtttcttta 3' Tm 61°C (SEQ ID NO:1666)

15 The PCR product was then digested with XhoI restriction enzyme, gel purified and then cloned into pPDM His, a modified pET28 vector with a His tag in frame, which had been digested with Eco72I and XhoI restriction enzymes. The correct construct was confirmed by DNA sequence analysis and then transformed into BL21 (DE3) pLys S and BL21 (DE3) CodonPlus RIL expression hosts. The protein sequence of expressed recombinant L548S is
20 shown in SEQ ID NO:1667, and the DNA sequence of expressed recombinant L7548S is shown in SEQ ID NO:1668.

EXAMPLE 4

ADDITIONAL ANALYSES OF LUNG CHIP 5

25

SQL1, SCL1, SCL3 AND SCL4 LIBRARIES

Additional analyses were performed on lung chip 5 using a criteria of greater than or equal to 2-fold over-expression in tumor probe groups versus normal tissues and an average expression in normal tissues of less than or equal to 0.2. This resulted in the
30 identification of 109 non-redundant clones that are over-expressed in lung carcinomas. As

summarized in the Table 11 below, 19 cDNA clones were recovered from the lung squamous cell carcinoma subtracted library SQL1, 9 cDNA clones were recovered from the small cell lung carcinoma library SCL1, and 81 cDNA clones were recovered from the small cell lung carcinoma libraries SCL3 and SCL4.

5

Table 11

SEQ ID NO:	Seq. Ref.	Element (384)	Element (96)	Ratio	Mean Signal 1	Mean Signal 2	Library
1680	58456	p0003r03c13	R0001 E7	3.09	0.424	0.137	SQL1
1681	58458	p0003r03c10	R0001 F5	2.31	0.408	0.176	SQL1
1682	58462	p0003r04c16	R0001 H8	2.22	0.257	0.116	SQL1
1683	58469	p0003r07c12	R0002 F6	2.1	0.289	0.138	SQL1
1684	58470	p0003r09c21	R0003 A11	2.55	0.493	0.194	SQL1
1685	58482	p0003r12c19	R0003 G10	2.16	0.36	0.167	SQL1
1686	58485	p0003r12c10	R0003 H5	2.48	0.273	0.11	SQL1
1687	58501	p0004r04c23	R0005 G12	2.04	0.26	0.128	SQL1
1688	58502	p0004r04c03	R0005 G2	2.17	0.289	0.133	SQL1
1689	58505	p0004r05c23	R0006 A12	3.08	0.454	0.148	SQL1
1690	58507	p0004r06c11	R0006 C6	3.22	0.49	0.152	SQL1
1691	58509	p0004r07c15	R0006 E8	3.26	0.421	0.129	SQL1
1692	58512	p0004r09c03	R0007 A2	3.16	0.559	0.177	SQL1
1693	58527	p0004r12c22	R0007 H11	2.03	0.278	0.137	SQL1
1694	58529	p0004r14c09	R0008 C5	2.26	0.45	0.199	SQL1
1695	58531	p0004r16c01	R0008 G1	2.84	0.387	0.136	SQL1
1696	58537	p0005r02c08	R0009 D4	2.03	0.355	0.175	SQL1
1697	58539	p0005r03c08	R0009 F4	2.34	0.42	0.18	SQL1
1698	58545	p0005r07c21	R0010 E11	2.96	0.361	0.122	SQL1
1699	59319	p0005r10c04	R0011 D2	3.1	0.478	0.154	SCL1
1700	59322	p0005r12c01	R0011 G1	2.16	0.255	0.118	SCL1
1701	59348	p0006r11c12	R0015 F6	2.33	0.269	0.116	SCL1
1702	59350	p0006r14c13	R0016 C7	2.41	0.447	0.185	SCL1
1703	59363	p0007r02c16	R0017 D8	2.12	0.421	0.199	SCL1
1704	59365	p0007r03c20	R0017 F10	3.07	0.584	0.19	SCL1
1705	59370	p0007r04c10	R0017 H5	2.06	0.284	0.138	SCL1
1706	59373	p0007r05c23	R0018 A12	2.95	0.472	0.16	SCL1
1707	59376	p0007r06c02	R0018 D1	2.13	0.246	0.116	SCL1
1708	61050	p0011r02c10	R0033 D5	2.23	0.306	0.137	SCL3/4
1709	61051	p0011r03c23	R0033 E12	2.9	0.298	0.103	SCL3/4
1710	61052	p0011r03c08	R0033 F4	2.18	0.265	0.122	SCL3/4
1711	61054	p0011r03c16	R0033 F8	2.11	0.415	0.197	SCL3/4
1712	61056	p0011r04c13	R0033 G7	2.73	0.314	0.115	SCL3/4

SEQ ID NO:	Seq. Ref.	Element (384)	Element (96)	Ratio	Mean Signal 1	Mean Signal 2	Library
1713	61057	p0011r04c10	R0033 H5	2.45	0.463	0.189	SCL3/4
1714	61060	p0011r05c11	R0034 A6	3.28	0.536	0.164	SCL3/4
1715	61062	p0011r06c21	R0034 C11	2.73	0.526	0.192	SCL3/4
1716	61063	p0011r06c05	R0034 C3	3.61	0.513	0.142	SCL3/4
1717	61064	p0011r06c04	R0034 D2	2.58	0.477	0.185	SCL3/4
1718	61065	p0011r06c14	R0034 D7	4.91	0.55	0.112	SCL3/4
1719	61066	p0011r06c18	R0034 D9	2.38	0.285	0.12	SCL3/4
1720	61069	p0011r07c16	R0034 F8	2.25	0.426	0.189	SCL3/4
1721	61070	p0011r08c21	R0034 G11	2	0.234	0.117	SCL3/4
1722	61071	p0011r08c03	R0034 G2	2.76	0.321	0.116	SCL3/4
1723	61074	p0011r08c16	R0034 H8	3.02	0.399	0.132	SCL3/4
1724	61075	p0011r09c05	R0035 A3	3.83	0.498	0.13	SCL3/4
1725	61077	p0011r10c21	R0035 C11	2.12	0.306	0.144	SCL3/4
1726	61079	p0011r11c23	R0035 E12	2.04	0.22	0.108	SCL3/4
1727	61080	p0011r11c15	R0035 E8	2.76	0.299	0.108	SCL3/4
1728	61081	p0011r11c14	R0035 F7	2.37	0.303	0.128	SCL3/4
1729	61083	p0011r12c15	R0035 G8	2.29	0.351	0.153	SCL3/4
1730	61085	p0011r13c05	R0036 A3	2.62	0.43	0.164	SCL3/4
1731	61086	p0011r13c09	R0036 A5	2.53	0.398	0.157	SCL3/4
1732	61088	p0011r14c05	R0036 C3	4.26	0.702	0.165	SCL3/4
1733	61090	p0011r15c07	R0036 E4	3.16	0.429	0.136	SCL3/4
1734	61091	p0011r16c16	R0036 H8	3.54	0.634	0.179	SCL3/4
1735	61093	p0012r02c03	R0037 C2	2.2	0.265	0.121	SCL3/4
1736	61094	p0012r02c11	R0037 C6	15.17	1.79	0.118	SCL3/4
1737	61096	p0012r02c08	R0037 D4	2.44	0.27	0.111	SCL3/4
1738	61097	p0012r02c10	R0037 D5	4.52	0.81	0.179	SCL3/4
1739	61099	p0012r03c02	R0037 F1	3.34	0.39	0.117	SCL3/4
1740	61100	p0012r03c06	R0037 F3	2.03	0.233	0.114	SCL3/4
1741	61103	p0012r04c17	R0037 G9	2.48	0.413	0.167	SCL3/4
1742	61105	p0012r05c11	R0038 A6	3.26	0.501	0.154	SCL3/4
1743	61106	p0012r05c08	R0038 B4	2.46	0.354	0.144	SCL3/4
1744	61110	p0012r06c15	R0038 C8	2.18	0.41	0.188	SCL3/4
1745	61113	p0012r07c09	R0038 E5	2.47	0.376	0.152	SCL3/4
1746	61115	p0012r07c13	R0038 E7	2.57	0.483	0.188	SCL3/4
1747	61117	p0012r07c24	R0038 F12	2.18	0.235	0.108	SCL3/4
1748	61118	p0012r07c18	R0038 F9	4.44	0.605	0.136	SCL3/4
1749	61119	p0012r08c03	R0038 G2	2.97	0.35	0.118	SCL3/4
1750	61120	p0012r08c07	R0038 G4	2.23	0.323	0.144	SCL3/4
1751	61122	p0012r08c18	R0038 H9	2.23	0.373	0.168	SCL3/4
1752	61125	p0012r10c17	R0039 C9	2.1	0.22	0.105	SCL3/4
1753	61126	p0012r10c16	R0039 D8	2.47	0.345	0.14	SCL3/4
1754	61130	p0012r12c12	R0039 H6	2.66	0.282	0.106	SCL3/4
1755	61133	p0012r13c24	R0040 B12	2.25	0.27	0.12	SCL3/4
1756	61134	p0012r14c23	R0040 C12	2.23	0.228	0.102	SCL3/4

SEQ ID NO:	Seq. Ref.	Element (384)	Element (96)	Ratio	Mean Signal 1	Mean Signal 2	Library
1757	61135	p0012r14c03	R0040 C2	2.05	0.298	0.146	SCL3/4
1758	61137	p0012r14c02	R0040 D1	8.63	1.463	0.17	SCL3/4
1759	61139	p0012r14c14	R0040 D7	2.69	0.3	0.111	SCL3/4
1760	61143	p0012r16c02	R0040 H1	2.55	0.318	0.125	SCL3/4
1761	61144	p0012r16c18	R0040 H9	2.85	0.318	0.112	SCL3/4
1762	61148	p0013r02c19	R0041 C10	2.33	0.463	0.199	SCL3/4
1763	61151	p0013r02c03	R0041 C2	2.25	0.336	0.149	SCL3/4
1764	61155	p0013r04c07	R0041 G4	2.13	0.366	0.171	SCL3/4
1765	61156	p0013r05c05	R0042 A3	2.73	0.38	0.139	SCL3/4
1766	61159	p0013r06c24	R0042 D12	4.57	0.831	0.182	SCL3/4
1767	61160	p0013r07c19	R0042 E10	8.6	1.191	0.138	SCL3/4
1768	61163	p0013r07c18	R0042 F9	2.18	0.278	0.128	SCL3/4
1769	61167	p0013r10c12	R0043 D6	3.13	0.39	0.124	SCL3/4
1770	61172	p0013r12c03	R0043 G2	2	0.396	0.198	SCL3/4
1771	61173	p0013r12c07	R0043 G4	3.73	0.72	0.193	SCL3/4
1772	61176	p0013r13c04	R0044 B2	2.34	0.446	0.19	SCL3/4
1773	61177	p0013r14c01	R0044 C1	3.9	0.539	0.138	SCL3/4
1774	61183	p0013r15c14	R0044 F7	5.49	0.959	0.175	SCL3/4
1775	61185	p0013r16c24	R0044 H12	2.25	0.409	0.182	SCL3/4
1776	61188	p0014r01c07	R0045 A4	2.14	0.271	0.127	SCL3/4
1777	61192	p0014r02c19	R0045 C10	2.33	0.321	0.138	SCL3/4
1778	61198	p0014r04c24	R0045 H12	2.3	0.321	0.14	SCL3/4
1779	61201	p0014r06c22	R0046 D11	2.43	0.269	0.111	SCL3/4
1780	61202	p0014r06c08	R0046 D4	2.57	0.346	0.135	SCL3/4
1781	61204	p0014r07c07	R0046 E4	4.27	0.516	0.121	SCL3/4
1782	61206	p0014r07c12	R0046 F6	2.18	0.364	0.167	SCL3/4
1783	61210	p0015r09c02	R0051 B1	2.43	0.463	0.19	SCL3/4
1784	61212	p0015r10c15	R0051 C8	2.64	0.406	0.154	SCL3/4
1785	61216	p0015r11c16	R0051 F8	2.28	0.278	0.122	SCL3/4
1786	61225	p0015r14c12	R0052 D6	2.25	0.25	0.111	SCL3/4
1787	61226	p0015r14c14	R0052 D7	2.54	0.3	0.118	SCL3/4
1788	61227	p0015r16c18	R0052 H9	2.06	0.312	0.151	SCL3/4

The ratio of signal 1 to signal 2 in the table above provides a measure of the level of expression of the identified sequences in tumor versus normal tissues. For example, for SEQ ID NO:1669, the tumor-specific signal was 3.09 times that of the signal for the normal tissues tested; for SEQ ID NO:1670, the tumor-specific signal was 2.31 times that of the signal for normal tissues, etc.

EXAMPLE 5

REAL-TIME PCR ANALYSES OF LUNG TUMOR SEQUENCES

Real-time PCR was performed on a subset of the lung tumor sequences disclosed herein in order to further evaluate their expression profiles in various tumor and normal tissues. Briefly, quantitation of PCR product relies on the few cycles where the amount of DNA amplifies logarithmically from barely above the background to the plateau. Using continuous fluorescence monitoring, the threshold cycle number where DNA amplifies logarithmically is easily determined in each PCR reaction. There are two fluorescence detecting systems. One is based upon a double-strand DNA specific binding dye SYBR Green I dye. The other uses TaqMan probe containing a Reporter dye at the 5' end (FAM) and a Quencher dye at the 3' end (TAMRA) (Perkin Elmer/Applied Biosystems Division, Foster City, CA). Target-specific PCR amplification results in cleavage and release of the Reporter dye from the Quencher-containing probe by the nuclease activity of AmpliTaq Gold™ (Perkin Elmer/Applied Biosystems Division, Foster City, CA). Thus, fluorescence signal generated from released reporter dye is proportional to the amount of PCR product. Both detection methods have been found to generate comparable results. To compare the relative level of gene expression in multiple tissue samples, a panel of cDNAs is constructed using RNA from tissues and/or cell lines, and real-time PCR is performed using gene specific primers to quantify the copy number in each cDNA sample. Each cDNA sample is generally performed in duplicate and each reaction repeated in duplicated plates. The final Real-time PCR result is typically reported as an average of copy number of a gene of interest normalized against internal actin number in each cDNA sample. Real-time PCR reactions may be performed on a GeneAmp 5700 Detector using SYBR Green I dye or an ABI PRISM 7700 Detector using the TaqMan probe (Perkin Elmer/Applied Biosystems Division, Foster City, CA).

Results obtained from real-time PCR analysis of a number of lung tumor-specific sequences disclosed herein are summarized in the table below. In addition, extended cDNA sequences for many of these clones were obtained by searching public

sequence databases. The extended sequences, and the proteins encoded by those sequences, are identified by SEQ ID NO: in Table 12 below.

Table 12

5

Library	Clone Name	Clone No	SEQ ID NO:	Real-Time PCR Results	Extended cDNA Sequence	Encoded Polypeptide Sequence
SQL1	L972P	47988	1789	Overexpressed in 1/7 lung squamous tumor, 1/3 HN squamous tumor. Low or no expression in normal tissues.		
SQL1	L979P	48005	1790	Over-expressed in 2/7 squamous lung tumors, 1/3 HN squamous tumors, 1/2 adeno lung tumors. Low or no expression in normal tissues.	1791	1806
SQL1	L970P	49853	1269	Highly overexpressed in 1/7 lung squamous tumors and 1/3 HN squamous tumor. Low or no expression in normal tissues.		
SQL1	L981P	49865	1272	Over-expressed in 1/6 squamous lung and 1/3 HN squamous tumors. Low or no expression in normal tissues.		
SQL1	L980P	49826	1279	Over-expressed in 3/7 squamous lung tumors, 1/3 HN squamous tumors, 1/2 adeno lung tumors. Low or no expression in normal tissues.	1792	1807
SCL1	L973P	20631	117	Over-expressed in atypical carcinoid METs and adenocarcinoma. Expression in several normal tissues.	1793	1808
SCL1	L974P	20661	128	Over-expressed in primary small cell, squamous and adenocarcinomas. Expression observed in several normal tissues.	1794	1809
SCL1	L996P	50430	1442	Over-expressed in 2/2 Primary Small Cell, 6/6 Small Cell Cell Lines, 1/1 Atypical Carc. METs, 1/1 Adeno, 1/1 Squamous. Very low or no expression in normal tissues.	1795	1810
SCL3	L977P	26961	288	Over-expressed in 1/2 Primary Small Cell, 1/6 Small Cell-Cell Line, and 1/1 Carcinoid Mets. Very low or no expression in normal tissues.	1796	1811
SCL2	L978P	24928	1339	Over-expressed in 2/2 primary small cell, 3/6 small cell-cell lines, 1/1 carcinoid mets., adeno and squamous tumor pools; Low or no expression in normal tissues	1797	1812
SCL3/4	L984P	50507	1446	Highly expressed in 1/2 primary small cell tumors and 4/6 small cell tumor cell lines. Low or no expression in normal tissues.	1798	1813

Library	Clone Name	Clone No.	SEQ ID NO:	Real-Time PCR Results	Extended cDNA Sequence	Encoded Polypeptide Sequence
SCL3/4	L580S	50536	1449	Over-expressed in select small cell and squamous tumors. Some expression observed normal brain, bronchiol, soft palate and trachea.		
SCL3/4	L988P	50645	1531	Over-expressed in 1/2 Primary Small Cell, 1/2 Primary Small Cell, 6/6 Small Cell-Cell Lines, 1/1 Carcinoid Mets., Adeno & Squamous Tumor pool. Expressed in some normal tissues (brain, adrenal gland, salivary gland, trachea, thymus).	1799	1814
SCL3/4	L1423P	50625	1533	Over-expressed in 1/2 primary small cell, 5/6 small cell-cell lines, 1/1 carcinoid mets. Also expressed in normal brain and pituitary gland.	1800	1815
SCL3/4	L986P	50483	1490	Over-expressed in 1/2 primary small cell, 5/6 small cell-cell lines, 1/1 carcinoid mets., adeno and squamous tumor pool. Expressed in normal brain, pituitary gland and spinal cord.		
SCL3/4	L987P	50560	1527	Over-expression in 1/2 Primary Small Cell, 6/6 Small Cell-Cell Lines, 1/1 Carcinoid Mets. Expression in normal pituitary and adrenal glands.	1801	1816-1818
SCL3/4	L1424P	50639	1547	Over-expression in 1/2 Primary Small Cell and 1/1 Carcinoid Mets. Low or no expression in normal tissues.		
MLN1	L997P	26749	730	Over-expression in 1/1 atypical carcinoid METs No expression in normal tissues.		
MLN1	L999P	26752	733	Over-expressed in 2/2 Primary Small Cell, 6/6 Small Cell Cell Lines, 1/1 Atypical Carc. METs Expression in several normal tissues.		
MLN1	L1400P	26529	405	Over-expressed in 2/6 Small Cell Cell Lines and 2/2 Primary Small Cell. Moderate to low expression in several normal tissues		
MLN1	L998P	27699	468	Over-expression in 1/1 Atypical Carcinoid METs. Low expression in normal tissues.	1802	1819
LAP1	L1425P	59303	949	Over-expressed in 4/7 squamous tumors, 1/2 adenocarcinoma tumors and in a pool of six small cell lung carcinomas. Moderate to high expression observed in normal brain, kidney and skeletal muscle.	1803	1820
LAP1	L1426P	59314	1156	Highly overexpressed in one lung squamous tumor and one HN squamous tumor. Very low or no expression observed in normal tissues.	1804	1821
LAP1	L1427P	59298	921	Highly over-expressed in 3/12 adenocarcinoma tumors. Very low or no expression in normal tissues	1805	1822

Library	Clone Name	Clone No.	SEQ ID NO:	Real-Time PCR Results	Extended cDNA Sequence	Encoded Polypeptide Sequence
LAP1	L1428P	59316	1180	Over-expressed in 4/12 adenocarcinoma tumors and lower level expression in several other adenocarcinoma tumors. Very low or no expression in normal tissues.		

110121.478C17

EXAMPLE 6IDENTIFICATION OF CD4 IMMUNOGENIC T CELL EPITOPES
DERIVED FROM LUNG TUMOR ANTIGENS

5

CD4 T cell lines specific for the antigen L548S (SEQ ID NO:789) were generated as follows.

A series of overlapping 20-mer peptides were synthesized that spanned the entire L548S sequence (SEQ ID NO:1834-1856, respectively). For priming, peptides were
10 combined into pools of 4-5 peptides and cultured at 2 micrograms/ml with dendritic cells and purified CD4+ T cells in 96 well U-bottomed plates. One hundred cultures were generated for each peptide pool. Cultures were restimulated weekly with fresh dendritic cells loaded with peptide pools. Following a total of 3 stimulation cycles, cells were rested for an additional week and tested for specificity to antigen presenting cells (APC) pulsed
15 with peptide pools using interferon-gamma ELISA and proliferation assays. For these assays, adherent monocytes loaded with either the relevant peptide pool, recombinant L548S or an irrelevant peptide were used as antigen presenting cells (APC). As shown in Table 13, below, a number of CD4 T cell lines demonstrated reactivity with the priming peptides as well as recombinant L548S protein. These lines were further expanded to be
20 tested for recognition of individual peptides from the pools, as well as for recognition of recombinant L548S.

The dominant reactivity of these lines appeared to be with peptide 21 (SEQ ID NO:1854), which corresponds to amino acids 161-180 of L548S.

Table 13

Stimulation		Proliferation (CPM)			Stimulation Index	
Peptides	Positive cell lines	No antigen	Peptides	L548S	Peptides	L548S
p1-5	B1	700	14891	8791	21	13
	B2	1135	48724	53944	42	47
	G2	1227	7609	3193	6	3
p6-10						
p11-15	E2	8315	33723	13391	4	2
	E4	22097	100040	44171	5	2
p16-19						
p20-23	E3	3937	45367	15524	11	4
	F4	2648	130947	12927	49	5

5

EXAMPLE 7

DETECTION OF ANTIBODIES AGAINST LUNG TUMOR ANTIGENS
IN PATIENT SERA

10 Antibodies specific for the lung tumor antigens L548S (SEQ ID NO:789), and L552S (SEQ ID NO:809) were shown to be present in effusion fluid or sera of lung cancer patients but not in normal donors. More specifically, the presence of antibodies against L548S, L551S (SEQ ID NO:827) and L552S in effusion fluid obtained from lung cancer patients and in sera from normal donors was examined by ELISA using recombinant

15 proteins and HRP-conjugated anti-human Ig. Briefly, each protein (100 ng) was coated in a 96-well plate at pH 9.5. In parallel, BSA (bovine serum albumin) was also coated as a control protein. The signals ([S], absorbance measured at 405 nm) against BSA ([N]) were determined. The results of these studies are shown in Table 14, wherein - represents [S]/[N] < 2; +/- represents [S]/[N] >2; ++ represents [S]/[N] >3; and +++ represents

20 [S]/[N] >5.

Table 14

Detection of Antibodies against Lung Tumor Antigens

	L548S	L551S	L552S
Effusion fluid			
#1	+/-	-	-
#2	-	-	-
#3	-	-	+++
#4	-	-	-
#5	+/-	-	-
#7	-	-	-
#8	-	-	+/-
#10	-	-	+/-
#11	-	-	-
#12	=	-	+/-
#13	-	-	-
#14	+/-	-	+/-
#15	-	-	-
#17	-	-	++
#18	-	-	-
#19	-	-	++
#20	-	-	-
Normal sera			
#21	-	-	-
#22	-	-	-
#23	-	-	-
#24	-	-	-
#25	-	-	-

- 5 Using Western blot analyses, antibodies against L552S were found to present in 1 out of 4 samples of effusion fluid from lung cancer patients, with no L552S antibodies being detected in the three samples of normal sera tested.

EXAMPLE 8

FUSION PROTEINS OF LUNG TUMOR ANTIGENS

Fusion proteins of full-length Ra12 with either L801P ORF4 (SEQ ID NO:1670) or L801P ORF5 (SEQ ID NO:1671) were prepared and expressed as single recombinant proteins in *E. coli* as follows.

The cDNA for ORF4 of L801P was obtained by PCR with a cDNA for the full length L801P and the primers of SEQ ID NO:1857 and 1858. The cDNA for ORF5 of L801P was obtained by PCR with a cDNA for the full length L801P and the primers of SEQ ID NO:1859 and 1860. The PCR products with expected size were recovered from agarose gel, digested with restriction enzymes EcoRI and XhoI, and cloned into the corresponding sites in the expression vector pCRX1 for subsequent expression in *E. coli*. For the fusion of Ra12 with ORF4, the best expression was obtained in HMS174(DE3)pLysS in 2xYS media, with recombinant protein being induced using IPTG at 37 °C for approximately 3 hours. For the fusion of Ra12 with ORF5, the best expression was obtained in HMS174(DE3)pLysS in 2xYS media, again with recombinant protein being induced using IPTG at 37 °C for approximately 3 hours. The plasmids used for the fusion protein production were confirmed by DNA sequencing. The determined cDNA sequences for the ORF4 and ORF5 fusions are provided in SEQ ID NO:1861 and 1862, respectively, with the corresponding amino acid sequences being provided in SEQ ID NO:1863 and 1864, respectively.

EXAMPLE 9

CLONING OF CDNA ENCODING FULL-LENGTH L984P

25

The example illustrates the isolation of cDNA sequences encoding L984P by PCR amplification from four separate cDNA sources. Briefly, an earlier isolated cDNA sequence of clone L984P was identified as having homology to a DNA sequence that encodes human achaete-scute homolog 1 (ASH1). Gene specific primers were made using

the sequence information present in the public domain for ASH1 (genbank acc. NM-004316). Using these gene specific primers in PCR amplification, L984P was cloned from four separate cDNA sources. The four cDNA sources were a small cell lung carcinoma primary tumor sample (RNA Id. 573A), a METs neuroendocrine atypical carcinoid sample (RNA Id. 512A), and two small cell lung carcinoma cell lines (cell-line Id. NCI H128 and DMS79). The determined cDNA sequences for these four clones are provided in SEQ ID NO:1865-1868, respectively. The coding region of the cDNA contains a repeat of the triplet CAG that exhibits polymorphism in the human genomic DNA. This polymorphism can be observed in L984P cloned from the four different sources as well as from the cDNA of ASH1 and that derived from the human chromosome 4. The cDNA of ASH1 (genbank Acc. NM-006688) deposited in the genbank database contains 14 copies of the triple CAG, whereas the cDNA sequence derived from the human chromosome 4 sequence (genbank Acc. XM_006688) contains 12 copies of the triplet. The cDNA cloned from 573A and 512A both contain 12 copies of the CAG triplet. The cDNA cloned from the small cell lung carcinoma cell line DMS79 contains 13 copies of the triplet CAG, while the cDNA cloned from the small cell lung carcinoma cell line NCI H128 contains only 10 copies of the triplet CAG. As the polymorphism is present in the coding region, this results in polymorphisms in the protein sequences as well (*see*, SEQ ID NO:1869-1872).

EXAMPLE 10

CLONING OF CDNA ENCODING FULL-LENGTH L985P

As previously disclosed in Example 1C, a search of the public databases using the sequence for contig 139 (SEQ ID NO:1467, also known as L985P) as the query was conducted and showed that this sequence had homology to the cDNA (SEQ ID NO:1676) encoding the cell surface immunomodulator-2 (CSIMM-2, sequence obtained from the Geneseq database). The full-length sequence of the clone was obtained by screening a small cell lung carcinoma cDNA library with a radioactively labeled probe of the original cloned sequence (SEQ ID NO:1467). Approximately 500,000 clones from the

cDNA library were screened and 2 independent clones containing cDNA insert of 1.35 kb were isolated. The full-length cDNA sequence is provided in SEQ ID NO:1873 and the encoded amino acid sequence is provided in SEQ ID NO:1874. Surprisingly, an alignment of the isolated full-length cDNA sequence of L985P (SEQ ID NO:1873) with the Geneseq database sequence for CSIMM-2 showed that L985P differs from the GeneSeq database sequence by one nucleotide. This nucleotide difference results in a change of one amino acid residue at position 119 (G to E) of SEQ ID NO:1874.

The predicted protein structure and sequence of L985P indicates that it is a member of the recently described MS4A (membrane-spanning 4-domain, subfamily A) gene family (Liang and Tedder, 2001 Genomics 72:119-127). The MS4A gene family currently consists of at least 21 distinct human and mouse proteins of which nine members are from humans. These include CD20, FcεRIβ, HTm4, MS4A4A, MS4A5, MS4A6A, MS4A7, MS4A8B (same as L985P) and MS4A12. The MS4A family members are cell surface expressed proteins containing four transmembrane spanning domains, with N- and C-terminal regions facing the cytoplasmic side of the cells. The human MS4A family members exhibit 20-40% overall homology at the protein level, which is confined mostly to the transmembrane domains. The transmembrane domains of L985P share the highest homology to CD20 with approximately 40% identity and 60% similarity between the two protein sequences in this region. The MS4A family members demonstrate a broad tissue distribution with expression observed in diverse cell types in hematopoietic and nonhematopoietic tissues. However, expression of some MS4A family members is highly restricted to a particular cell type, such as CD20, which is only expressed on B-cells. As also mentioned in Example 1C, L985P (MS4A8B) is over-expressed in small cell lung carcinoma (SCLC) as determined by quantitative real-time PCR. Low level expression is seen in some normal tissues including lung, pituitary gland, stomach, colon, and trachea, while expression in other normal tissues checked was negligible or undetectable (*see*, Example 1C).

The physiological function for most of the MS4A family members remains to be elucidated. Previous studies have shown that CD20 is functionally important for the

regulation of cell growth and differentiation, and signal transduction in B-cells. There is also evidence that CD20 may serve as a calcium channel by forming a homo- or heterotetrameric complex. FcεRIβ is part of a tetrameric receptor complex, which mediates interaction with IgE-bound antigens that lead to cellular responses such as the degranulation of mast cells. Because of the sequence and structural homologies between the MS4A family members, it is highly likely that they will share overlapping functional properties.

Some of the MS4A family members have been found to be associated with cancer. CD20 is expressed in more than 90% of B-cell non-Hodgkin lymphomas, which has made it an ideal target for immunotherapeutic approaches for the treatment of B-cell malignancies. Anti-CD20 monoclonal antibodies have been used with high success in the treatment of non-Hodgkin-lymphomas in both naïve and radiolabeled forms. The anti-CD20 MAbs have been shown to exert their antitumor effects through several pathways, which include complement-mediated cytotoxicity, antibody-mediated cellular cytotoxicity and antibody-mediated cell cycle arrest and apoptosis. Analysis of the human EST databases indicates that other members of the MS4A gene family including MS4A4A, A6A, A7 and A8B are also expressed in various cancers including lung, breast, pancreas, colon, ovary, kidney and brain. By comparison to CD20, one or more of the MS4A family genes would be used as targets for immunotherapeutic approaches for the treatment of hematopoietic and nonhematopoietic malignancies.

The MS4A protein family members are structurally similar to other membrane protein families with four transmembrane domains. These include the Tetraspanin protein family (TM4SF) and the GABA-A receptor protein family. Tetraspanins are associated with cancer and may play a direct role in controlling tumor progression. Although CD9 expression will positively influence B cell migration, CD9 overexpression suppresses motility and metastasis in carcinoma cells and there is an inverse correlation with metastasis in melanoma. However, CD9 is also expressed on 90% of non-T cell acute lymphoblastic leukemia cells and 50% of chronic lymphocytic leukemias. A recent study using RT-PCR analysis of tetraspanin expression in Burkitt lymphoma cell

lines found that 90% of the lines express CD53, CD81, CD63, CD82 and SAS at high levels. CD151/PETA3 is an effector of metastasis and cell migration and MAbs that block this activity have been developed. Similarly, overexpression of the tetraspanin CO-029/D6.1 will increase the metastatic potential of cell lines. The tetraspanins control a

5 diverse set of biological functions that can be regulated by MAbs. The functions of the tetraspanins, in general, can be grouped into actions that affect cell activation and proliferation, as well as adhesion and motility. These functions tend to be carried out by their association with integrins. The functional activity of tetraspanins can be modulated with MAbs in such a way as to control cell proliferation. For example, CD81/TAPA-1 is

10 associated with B cell activation and increased proliferation, an activity that can be blocked with MAbs. MAbs with anti-proliferative activity have been generated to the tetraspanin family member CO-029/D6.1..

As mentioned above, L985P is specifically over-expressed in small cell lung carcinomas. This fact and a comparison to CD20, tetraspanins and Her2 whose over-

15 expression in cancers makes them effective cancer therapeutic targets, indicate that L985P may be a good target for immunotherapeutic approaches for the treatment of small cell lung carcinomas.

To facilitate the generation, purification, and evaluation of MAb against L985P, MAbs against the entire deduced amino acid sequence of the L985P protein,

20 peptides derived from L985P or chemically produced (synthetic) L985P peptides will be used. Also, one can use MAbs raised against chimeric forms of L985P protein molecule fused to Ra12 protein, either the long form (Ra12- which is the first 128 amino acids of Ra12) and/or the short form (Ra12S), or fused to a polyhistidine peptide or any combination of these molecules. Provided are the predicted cDNA and amino acid

25 sequences for the his-tagged L985P-Ra12 fusion molecules: Ra12-L985P_cDNA (SEQ ID NO:1875), Ra12-L985P_Protein (SEQ ID NO:1876), Ra12S-L985P_cDNA (SEQ ID NO:1877) and Ra12S-L985P_Protein (SEQ ID NO:1878); and the L985P derived peptides: his-tagged Ra12S-L985PEx_cDNA (SEQ ID NO:1879), his-tagged Ra12S-

L985PEx_Protein (SEQ ID NO:1880), L985P_Extracellular_Loop-2_cDNA (SEQ ID NO:1881) and L985P_Extracellular_Loop-2_Peptide (SEQ ID NO:1882).

EXAMPLE 11

5 EXPRESSION IN *E. COLI* OF A HIS-TAGGED RA12-L985P FUSION PROTEIN

This example sets forth a specific embodiment of a fusion between Ra12 and L985P and its expression in *E. coli*. A his-tagged fusion protein of the long-form of Ra12 and all but the first three amino acid residues of L985P was expressed as a single
10 recombinant protein in *E. coli*. The long-form of Ra12 was modified from the original sequence of amino acid residues 192-323 of MTA32A in that a putative thrombin cleavage site was replaced with a HindIII restriction site. The L985P was fused downstream of the Ra12 sequence in a pCRX1 vector. The Ra12 sequence was cloned downstream of the RBS. As a result, a his-tagged fusion protein is produced when the recombinant vector is
15 expressed in *E. coli*. The sequence for the fusion of the long-form of Ra12 and L985P was confirmed by DNA sequencing. The determined cDNA sequence is provided in SEQ ID NO:1875 as this sequence is the same as that predicted in Example 10.

EXAMPLE 12

20

CLONING OF CDNA ENCODING FULL-LENGTH L1428P

As previously disclosed in Example 5, real-time PCR was performed on a subset of the lung tumor sequences disclosed herein in order to further evaluate their
25 expression profiles in various tumor and normal tissues. The results are provided above in Table 12. One of the sequences analyzed was from clone #59316 (SEQ ID NO:1180, L1428P) and was shown to be expressed in a subset of lung adenocarcinomas.

Further studies have isolated the full-length cDNA for the cloned sequence of #59316 (SEQ ID NO:1180, L1428P). In order to determine the transcript size of the

gene, a multiple tissue Northern blot was probed with the radioactively labeled original cloned sequence (SEQ ID NO:1180). The Northern blot included about 20 ug of total RNA from lung adenocarcinoma and normal tissues samples. Visual analysis of the exposed film revealed a single transcript of approximately 6.5 - 7.0 kb. The full-length sequence of the clone was obtained by screening a lung adenocarcinoma primary tumor cDNA library with a radioactively labeled probe of the original cloned sequence (SEQ ID NO:1180). Approximately 500,000 clones from the cDNA library were screened and 5 independent clones containing a cDNA insert of 6.8 kb were isolated. This insert size is similar to the size estimated by Northern-blot analysis. The full-length sequence is provided in SEQ ID NO:1883. Although no distinct ORF could be identified, seven potential ORFs can be predicted and the amino acid sequences of these potential ORFs, designated L1428P_ORF1 to ORF7, are provided in SEQ ID NO:1884-1890, respectively. The expression of full-length L1428P was re-analyzed by real-time PCR as set forth in Example 5 on extended cDNA panels for both lung adenocarcinomas and squamous cell carcinomas. The lung adenocarcinoma extended panel real-time results again confirm the expression of L1428P in adenocarcinoma with about 40-50% of the adenocarcinoma samples showing varying levels of expression. The real-time results from the squamous cell carcinoma extended panel shows that L1428P is also expressed in lung squamous cell carcinoma. However, the expression was in fewer lung squamous cell carcinoma samples and at a lower level.

20

EXAMPLE 13

REAL-TIME PCR ANALYSIS OF CDNA SEQUENCES WHICH ARE OVER-EXPRESSED IN LUNG TUMORS AS SHOWN BY MICROARRAY

The following clones listed in Table 15 were shown previously to be over-expressed in lung tumors by microarray analysis (*see*, Example 2C and Example 6, Table 11). The results of this microarray analysis are summarized below in Table 15.

25

Table 15

SEQ ID NO:	CLONE ID #	Ratio	Mean Signal 1	Mean Signal 2
1383	50096 (contig 156)	16.04	1.895	0.118
1560	54454 (contig 234)	2.86	0.2	0.07
1561	54463 (contig 235)	3.31	0.248	0.075
1707	59376	3.23	0.734	0.227
1733	61090	2.68	0.364	0.136
1735	61093	5.7	0.687	0.121
1758	61137	8.63	1.463	0.17
1761	61144	7.02	0.783	0.112
1766	61159	4.57	0.831	0.182
1771	61173	5.57	1.075	0.193
1775	61185	5.29	0.962	0.182
1786	61225	3.41	0.379	0.111

Real-time PCR analysis was performed on these sequences on a small cell lung carcinoma panel as described in Example 7. The results obtained from the real-time PCR analysis are summarized in Table 16.

Table 16

SEQ ID NO:	Real-Time PCR Results
1383	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinoma and 6/6 SCLC cell lines. Some expression is also seen in normal brain, pituitary gland, spinal cord, and thymus.
1560	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas and 3/6 SCLC cell lines. Expression is also seen in normal brain and pituitary gland.
1561	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 6/6 SCLC cell lines, 1/1 atypical carcinoid metastases, adenocarcinoma, and squamous cell carcinoma pools. Expression is also seen in normal bone marrow, lymph node, thymus, and at lower levels in other normal tissues.
1707	On the SCLC panel, this gene is overexpressed in 1/6 SCLC cell lines and 0/2 primary small cell carcinomas. It is also expressed in normal stomach and at lower levels in normal brain, salivary gland, and trachea.
1733	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 6/6 SCLC cell lines, 1/1 atypical carcinoid metastases, adenocarcinoma, and squamous cell carcinoma pools. Some expression is also seen in normal bone marrow, lymph node, thymus, and at lower levels in other normal tissues.
1735	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 5/6 SCLC cell lines, 0/1 atypical carcinoid metastases, and adenocarcinoma pool. Lower level expression is also seen in normal bone marrow and skeletal muscle.
1758	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 6/6 SCLC cell lines, 0/1 atypical carcinoid metastases, adenocarcinoma, and squamous cell carcinoma pools. Some expression is also seen in normal bone marrow, thyroid gland, and trachea.
1761	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 6/6 SCLC cell lines, 0/1 atypical carcinoid metastases, adenocarcinoma, and squamous cell carcinoma pools. Expression is also seen in normal bone marrow.

SEQ ID NO:	Real-Time PCR Results
1766	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 6/6 SCLC cell lines, and 0/1 atypical carcinoid metastases. Expression is also seen in normal pituitary gland, adrenal gland, bone marrow, thymus, salivary gland, and at lower levels in a variety of other normal tissues.
1771	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 6/6 SCLC cell lines, 0/1 atypical carcinoid metastases, and squamous cell carcinoma pools. Some expression is also seen in normal pituitary gland.
1775	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, adenocarcinoma, and squamous cell carcinoma pools. No expression is observed in the SCLC cell lines and the atypical carcinoid metastases. Some normal tissue expression is seen in liver, stomach, thyroid gland, lymph node, and thymus.
1776	On the SCLC panel, this gene is overexpressed in 2/2 primary small cell carcinomas, 6/6 SCLC cell lines, adenocarcinoma, and squamous cell carcinoma pools. Some expression is also seen in normal bone marrow, pituitary gland, stomach, trachea, and thymus.

These sequences were then compared to known sequences in the available databases (Genbank, GeneSeq, huEST, etc.). Nine of these sequences showed some degree of similarity to known sequences in the available databases. The results of these nine hits are summarized in Table 17 along with providing sequence listing identifiers for the DNA sequences and the respective encoded amino acid sequences (where available) of these matches. SEQ ID NO:1561 and 1786 showed no significant similarity to any known sequences.

Table 17

SEQ ID NO:	GenBank Database Hit	GeneSeq. Hit	cDNA Seq. of Hit	Amino Acid Seq. Of Hit (If Known)
1383	Pr22 Protein/Stathmin/Oncoprotein 18	A16376, A08801, A01633	1891	1901
1560	CDNA DKFZp564O163	--	1892	--
1561	Novel	Human secreted protein 5' EST. (C30107)	--	--
1707	SOX21 (AF107044)	Human secreted protein gene 7 cloone HE8CV18. (X27317)	1893	1902
1733	KIAA0166 gene (D79988)	Human gene signature HUMGS08725. (T26483)	1894	1903
1735	Ubiquitin-conjugating enzyme E2 (AF160215)	DNA encoding human ubiquitin-like conjugating protein (UBCLE). (X81676)	1895	1904
1758	Pituitary tumor transforming gene protein 1 (AF095287)	Z97293, X89295, V36964, V63198, Z97292, C00858, V88346, Z80287	1896	1905
1761	Novel	Kidney injury associated molecule HW051 cDNA clone. (V80605)	1897	1906
1766	Cyclin-dependent kinase inhibitor p18 (CDKN2C) (AF041248)	Cyclin-dependent kinase (CDK6) inhibiting protein. (T10925; T31456)	1898	1907
1771	CDK4-inhibitor (p16-INK4) (L27211)	Multiple Matches	1899	1908
1775	Monokine induced by gamma interferon (MIG) (NM 002416)	Monokine induced by gammer-interferon. (X14998; Z26088)	1900	1909
1776	Novel	--	--	--

Further studies have resulted in isolation of the full-length cDNA sequence for the cloned sequence of clone #61093 (SEQ ID NO:1735, L1437P). In order to determine the transcript size of the gene, a multiple tissue Northern blot was probed with the radioactively labeled original cloned sequence (SEQ ID NO:1735). The Northern blot

5 included about 20 ug of total RNA from small cell lung carcinoma and normal tissues samples. Visual analysis of the exposed film revealed a single transcript of approximately 1.2 kb. The full-length sequence was obtained by screening a small cell lung carcinoma tumor cDNA library with the radioactively labeled probe of the original cloned sequence (SEQ ID NO:1735). Approximately 120,000 clones from the cDNA library were screened

10 and 2 independent clones containing a cDNA insert of 931 bases were isolated. The inserts are similar in size to that estimated by Northern Blot analysis. The full-length cDNA sequence is provided in SEQ ID NO:1910. It was discovered that there was one nucleotide difference between the full-length cDNA and a previously published sequence. However, this nucleotide change does not result in a change in the deduced amino acid sequence.

15 The deduced amino acid sequence encoded by the full-length cDNA is the same as already provided in SEQ ID NO:1904, and confirms earlier predictions that this cDNA encodes a known protein, ubiquitin-conjugated enzyme E2 (AF160215, SEQ ID NO:1904). SEQ ID NO:1910 (L1437P) was shown to be over-expressed in lung small cell lung carcinoma by microarray, real-time PCR and Northern Blot analysis.

20

EXAMPLE 14

EXPRESSION IN *E. COLI* OF A L548S HIS TAG FUSION PROTEIN

25 PCR was performed on the L548S coding region with the following primers:

Forward primer PDM-433 5' gctaaaggtgacccaagaaaccaaag 3' (SEQ ID NO:1911) Tm 60°C.

Reverse primer PDM-438 5' ctattaactcgaggagacagataaacagtttcttta 3' (SEQ ID NO:1912) TM 61°C.

The PCR conditions were as follows:

10µl 10X Pfu buffer

1.0µl 10mM dNTPs

2.0µl 10µM each primer

5 83µl sterile water

1.5µl Pfu DNA polymerase (Stratagene, La Jolla, CA)

50ng DNA

96°C for 2 minutes, 96°C for 20 seconds, 61°C for 15 seconds, 72°C for 1 minute 30 seconds with 40 cycles and then 72°C for 4 minutes.

10 The PCR product was digested with XhoI restriction enzyme, gel purified and then cloned into pPDM His, a modified pET28 vector with a His tag in frame, which had been digested with Eco72I and XhoI restriction enzymes. The correct construct was confirmed by DNA sequence analysis and then transformed into BL21 CodonPlus (Stratagene, La Jolla, CA) and BL21 pLys S (Novagen, Madison, WI) cells for expression.

15 The amino acid sequence of expressed recombinant L548S is shown in SEQ ID NO:1913, and the DNA coding region sequence is shown in SEQ ID NO:1914.

EXAMPLE 15

20 EXPRESSION IN *E. COLI* OF A L551S HIS TAG FUSION PROTEIN

PCR was performed on the L551S coding region with the following primers:

Forward primer PDM-498 5' gtgacgatggaggagctgcgggagatgg 3' (SEQ ID NO:1915) Tm 67°C.

25 Reverse primer PDM-499 5' cgcctaactcgagtcactaacagctgggag 3' (SEQ ID NO:1916) TM 66°C.

The PCR conditions were as follows:

10µl 10X Pfu buffer

1.0µl 10mM dNTPs

2.0μl 10μM each primer

83μl sterile water

1.5μl Pfu DNA polymerase (Stratagene, La Jolla, CA)

50ng DNA

5 96°C for 2 minutes, 96°C for 20 seconds, 66°C for 15 seconds, 72°C for 2 minutes 20 seconds with 40 cycles and then 72°C for 4 minutes.

The PCR product was digested with XhoI restriction enzyme, gel purified and then cloned into pPDM His, a modified pET28 vector with a His tag in frame, which had been digested with Eco72I and XhoI restriction enzymes. The correct construct was confirmed by DNA sequence analysis and then transformed into BLR (DE3) pLys S and BLR (DE3) CodonPlus RP cells for expression.

The amino acid sequence of expressed recombinant L551S is shown in SEQ ID NO:1917, and the DNA coding region sequence is shown in SEQ ID NO:1918.

15

EXAMPLE 16

EXPRESSION IN *E. COLI* OF A L552S HIS TAG FUSION PROTEIN

PCR was performed on the L552S coding region with the following primers:

20 Forward primer PDM-479 5' cgggtgccacgcccatggaccttc 3' (SEQ ID NO:1919) Tm 64°C.

Reverse primer PDM-480 5' ctgagaattcattaaactgtggttgccttcacc 3' (SEQ ID NO:1920) TM 62°C.

The PCR conditions were as follows:

25

10μl 10X Pfu buffer

1.0μl 10mM dNTPs

2.0μl 10μM each primer

83μl sterile water

1.5μl Pfu DNA polymerase (Stratagene, La Jolla, CA)

50ng DNA

96°C for 2 minutes, 96°C for 20 seconds, 63°C for 15 seconds, 72°C for 1 minute with 40 cycles and then 72°C for 4 minutes.

The PCR product was digested with EcoRI restriction enzyme, gel purified
5 and then cloned into pPDM His, a modified pET28 vector with a His tag in frame, which had been digested with Eco72I and EcoRI restriction enzymes. The correct construct was confirmed by DNA sequence analysis and then transformed into BL21 CodonPlus (Stratagene, La Jolla, CA) cells for expression.

The amino acid sequence of expressed recombinant L552S is shown in SEQ
10 ID NO:1921, and the DNA coding region sequence is shown in SEQ ID NO:1922.

EXAMPLE 17

CLONING OF CDNA ENCODING FULL-LENGTH CLONES #19069 AND CLONE

15 #18965 OR #19002

Partial sequences of two lung antigens, clones #19069 (SEQ ID NO:90) and #18965 or #19002 (both SEQ ID NO:15), were previously provided. These partial sequences were used as a query to predict the full-length cDNA sequences for the isolated
20 cloned sequenced by searching the publicly available databases. The predicted full-length cDNA sequence for the isolated cloned sequence of SEQ ID NO:90 is provided in SEQ ID NO:1923. The predicted full-length cDNA sequence for the isolated cloned sequence of SEQ ID NO:15 is provided in SEQ ID NO:1924. The deduced amino acid sequences of the two antigens are provided in SEQ ID NO:1925 and 1926, respectively. These sequences
25 were then compared to known sequences in the GeneSeq database. Both sequences showed some degree of similarity to known sequences in the GeneSeq database. SEQ ID NO:1923 shows similarity to a lipophosphatic acid acyltransferase (GeneSeq Z25000 and Z65038) and SEQ ID NO:1924 shows similarity to a zinc/iron regulated transporter-like protein (Geneseq Z38333 and A14995).

EXAMPLE 18

EPI TOPE-MAPPING OF L552S-SPECIFIC ANTIBODIES

5 Peptides of candidate antigens can be used for the evaluation of antibody responses in both preclinical and clinical studies. These data allow one to further confirm the antibody response against a certain candidate antigen. Protein-based ELISA with and without competitive peptides and peptide-based ELISA can be used to evaluate these antibody responses. Peptide ELISA is especially useful since it can further exclude the
10 false positive of the antibody titer observed in protein-based ELISA as well as to provide the simplest assay system to test antibody responses to candidate antigens. In this example, data was obtained using L552S-peptides that show that individual cancer patients produce L552S-specific antibodies recognizing primarily the following three epitopes of L552S:

- 15 (1) aa21-35: GPRSGGAQAKLGCCW (SEQ ID NO:1927)
 (2) aa116-135: KVICKSCISQTPGINLDLGS (SED ID NO:1928)
 (3) aa141-160: IIPKEEHCKMPEAGEEQPV (SED ID NO:1929)

 In further studies, it was found that affinity-purified antibodies generated by SEQ ID NO:1929 can recognize the L552S protein, and occupy about 0.6% of the total
20 immunoglobulin kappa of a patient's lung plural effusion (LPE) fluid. It was also found that SEQ ID NO:1929 is the dominant epitope of the rabbit polyclonal antibodies specific for L552S protein.

25

EXAMPLE 19

L985P EXPRESSION

 For recombinant expression in mammalian cells, the full length L985P cDNA was subcloned into the mammalian expression vector pCEP4 (Invitrogen) with and

without a FLAG epitope tag. Both constructs were transfected into HEK293 cells (ATCC) using Lipofectamine 2000 reagent (Invitrogen). Western blot analysis was then performed on these transfected cells to determine if recombinant L985P was being transiently expressed.

5 Briefly the transfection was carried out as follows. HEK cells were plated at a density of 350,000 cells/well (6 well plate) in DMEM (Gibco) containing 10% FBS (Hyclone) and grown overnight. The following day, 2 μ l of Lipofectamine 2000 (Invitrogen) was added to 50 μ l of Optimem 1 (Invitrogen) containing no FBS and incubated for 5 minutes at room temperature. In a different set of tubes 50 μ l of Optimem 1
10 was mixed with 0.8 μ g of L985P (with and without FLAG)/plasmid DNA and the mixture was transferred to the Lipofectamine 2000/Optimem mix. The combined mixture was incubated for 20 minutes at room temperature and transferred to the HEK293 cells containing 0.5ml of DMEM 10% FBS. The Lipofectamine 2000/DNA mix was then added to the HEK293 cells and incubated for 16-24 hrs at 37°C with 7% CO₂. Cells were rinsed
15 with PBS then collected and pelleted by centrifugation.

For Western blot analysis, whole cell lysates were generated by incubating the cells in Triton-X100 containing lysis buffer for 30 minutes on ice. Lysates were then cleared by centrifugation at 15,000 rpm for 5 minutes at 4°C. Samples were diluted with SDS-PAGE loading buffer containing beta-mercaptoethanol, then boiled for 10 minutes
20 prior to loading on the SDS-PAGE gel. The protein was transferred to nitrocellulose and probed using a purified anti-L985P rabbit polyclonal sera at a dilution of 1:1000. The blot was revealed with a donkey anti-rabbit Ig coupled to HRP (Jackson ImmunoResearch) followed by incubation in ECL substrate. Results of the blot indicate that recombinant L985P (with and without the FLAG) was expressed in the HEK293 cells.

25

EXAMPLE 20

GENERATION OF POLYCLONAL ANTIBODIES TO LUNG TUMOR ANTIGENS

Three lung antigens, L548S (SEQ ID NO:789), L552S (SEQ ID NO:809) and L985P peptide #3482 (SEQ ID NO:1930), were expressed and purified for use in antibody generation.

L548S and L552S were expressed in an *E. coli* recombinant expression system and grown overnight in LB Broth with the appropriate antibiotics at 37°C in a shaking incubator. The next morning, 10 ml of the overnight culture was added to 500 ml of 2x YT with the appropriate antibiotics in a 2L-baffled Erlenmeyer flask. When the optical density of the culture reached 0.4-0.6 at 560 nanometers, the cells were induced with IPTG (1 mM). Four hours after induction with IPTG, the cells were harvested by centrifugation.

The cells were then washed with phosphate buffered saline and centrifuged again. The supernatant was discarded and the cells were either frozen for future use or immediately processed. Twenty milliliters of lysis buffer was added to the cell pellets and vortexed. To break open the *E. coli* cells, this mixture was then run through a french press at a pressure of 16,000 psi. The cells were then centrifuged again and the supernatant and pellet were checked by SDS-PAGE for the partitioning of the recombinant protein.

For proteins that localized to the cell pellet, the pellet was resuspended in 10 mM Tris pH 8.0, 1% CHAPS and the inclusion body pellet was washed and centrifuged again. This procedure was repeated twice more. The washed inclusion body pellet was solubilized with either 8M urea or 6M guanidine HCl containing 10 mM Tris pH 8.0 plus 10 mM imidazole. The solubilized protein was added to 5 ml of nickel-chelate resin (Qiagen) and incubated for 45 minutes to 1 hour at room temperature with continuous agitation.

After incubation, the resin and protein mixture was poured through a disposable column and the flow through was collected. The column was then washed with 10-20 column volumes of the solubilization buffer. The antigen was then eluted from the column using 8M urea, 10 mM Tris pH 8.0 and 300 mM imidazole and collected in 3 ml fractions. A SDS-PAGE gel was run to determine which fractions to pool for further purification.

As a final purification step, a strong anion exchange resin, in this case Hi-Prep Q (Biorad), was equilibrated with the appropriate buffer and the pooled fractions from above were loaded onto the column. Each antigen was eluted off the column with an increasing salt gradient. Fractions were collected as the column was run and another SDS-
 5 PAGE gel was run to determine which fractions from the column to pool.

The pooled fractions were dialyzed against 10 mM Tris pH 8.0. The release criteria were purity as determined by SDS-PAGE or HPLC, concentration as determined by Lowry assay or Amino Acid Analysis, identity as determined by amino terminal protein sequence, and endotoxin level was determined by the Limulus (LAL) assay. The proteins
 10 were then put in vials after filtration through a 0.22-micron filter and the antigens were frozen until needed for immunization.

The L985P peptide #3482 was synthesized and conjugated to KLH and frozen until needed for immunization.

The polyclonal antisera were generated using 400 micrograms of each lung
 15 antigen combined with 100 micrograms of muramyl dipeptide (MDP). An equal volume of Incomplete Freund's Adjuvant (IFA) was added and then mixed and injected subcutaneously (S.C.) into a rabbit. After four weeks, the rabbit was S.C. boosted with 200 micrograms of antigen mixed with an equal volume of IFA. Thereafter the rabbit was I.V. boosted with 100 micrograms of antigen. The animal was bled seven days following each
 20 boost. The blood was then incubated at 4°C for 12-24 hours followed by centrifugation to generate the sera.

The polyclonal antisera were characterized using 96 well plates coated with antigen and incubating with 50 microliters (typically 1 microgram/microliter) of the polyclonal antisera at 4°C for 20 hours.

25 250 microliters of BSA blocking buffer was added to the wells and incubated at room temperature for 2 hours. Plates were washed 6 times with PBS/0.1% Tween. The rabbit sera were diluted in PBS/0.1% Tween/0.1% BSA. 50 microliters of diluted sera was added to each well and incubated at room temperature for 30 minutes. The plates were washed as described above, and then 50 microliters of goat anti-rabbit

horseradish peroxidase (HRP) at a 1:10000 dilution was added and incubated at room temperature for 30 minutes.

The plates were washed as described above, and 100 microliters of TMB Microwell Peroxidase Substrate was added to each well. Following a 15-minute incubation
 5 in the dark at room temperature, the colorimetric reaction was stopped with 100 microliters of 1N H₂SO₄ and read immediately at 450 nm. All the polyclonal antibodies showed immunoreactivity to the appropriate antigen.

Tables 18-20 show the antibody reactivity of rabbit antisera in serial dilution to the three lung antigens, L548S, L552S and L985P peptide #3482. The first column
 10 shows the antibody dilutions. The columns "Pre-immune sera" indicate ELISA data for two experiments using pre-immune sera. These results are averaged in the fourth column. The columns "anti-L548S, L552S or #3482" indicate ELISA data for two experiments using sera from rabbits immunized as described above, using the respective antigen, referred to as either L548S, L552S or #3482 in the tables.

15

Table 18

Antibody dilution	Pre-immune sera (1)	Pre-immune sera (2)	Average	Anti-L548S (1)	Anti-L548S (2)	Average
1:1000	0.17	0.10	0.14	0.51	0.51	0.51
1:2000	0.12	0.09	0.11	0.30	0.30	0.30
1:4000	0.09	0.08	0.08	0.17	0.20	0.19
1:8000	0.09	0.07	0.08	0.12	0.13	0.12
1:16000	0.09	0.08	0.08	0.09	0.12	0.11
1:32000	0.09	0.08	0.09	0.08	0.09	0.09
1:64000	0.11	0.09	0.10	0.11	0.12	0.12
1:128000	0.10	0.08	0.09	0.08	0.09	0.08
1:256000	0.09	0.08	0.08	0.11	0.09	0.10
1:512000	0.11	0.09	0.10	0.08	0.08	0.08
1:1024000	0.10	0.08	0.09	0.08	0.10	0.09
1:2048000	0.10	0.09	0.09	0.08	0.09	0.09

Table 19

5

Antibody dilution	Pre-immune sera (1)	Pre-immune sera (2)	Average	Anti-L552S (1)	Anti-L552S (2)	Average
1:1000	0.08	0.19	0.13	2.14	2.03	2.08
1:2000	0.08	0.06	0.07	1.93	1.97	1.95
1:4000	0.07	0.06	0.06	1.81	1.82	1.82
1:8000	0.08	0.06	0.07	1.63	1.64	1.64
1:16000	0.06	0.05	0.05	1.47	1.29	1.38
1:32000	0.06	0.05	0.06	1.03	1.10	1.06
1:64000	0.06	0.06	0.06	0.73	0.69	0.71
1:128000	0.06	0.05	0.06	0.44	0.48	0.46
1:256000	0.06	0.06	0.06	0.26	0.25	0.26
1:512000	0.07	0.06	0.06	0.16	0.15	0.16
1:1024000	0.06	0.07	0.06	0.12	0.10	0.11
0.00	0.06	0.06	0.06	0.06	0.06	0.06

Table 20

Antibody dilution	Pre-immune sera (1)	Pre-immune sera (2)	Average	Anti-#3482 (1)	Anti-#3482 (2)	Average
1:1000	0.10	0.07	0.09	2.10	2.07	2.08
1:2000	0.07	0.07	0.07	1.80	1.84	1.82
1:4000	0.07	0.06	0.06	1.78	1.80	1.79
1:8000	0.06	0.06	0.06	1.94	1.72	1.83
1:16000	0.06	0.06	0.06	1.75	1.74	1.74
1:32000	0.06	0.06	0.06	1.42	1.47	1.44
1:64000	0.06	0.06	0.06	1.12	1.17	1.14
1:128000	0.06	0.06	0.06	0.79	0.87	0.83
1:256000	0.06	0.06	0.06	0.70	0.65	0.68
1:512000	0.06	0.06	0.06	0.41	0.41	0.41
1:1024000	0.06	0.06	0.06	0.25	0.25	0.25
0	0.06	0.06	0.06	0.06	0.06	0.06

5 Table 21 shows the Protein A purification of the antibodies to the lung antigen, L548S.

Table 21

Purified Antibody Concentration (µg/ml)	Pro A pure	Pro A pure	Average
3.0	2.20	2.12	2.16
1.5	2.09	2.01	2.05
0.75	1.95	1.93	1.94
0.38	1.83	1.85	1.84
0.188	1.68	1.68	1.68
0.094	1.36	1.38	1.37
0.047	1.02	1.06	1.04
0.0234	0.68	0.73	0.71
0.0177	0.40	0.42	0.41
0.0059	0.24	0.25	0.24
0.0029	0.17	0.15	0.16

Purified Antibody Concentration ($\mu\text{g/ml}$)	Pro A pure	Pro A pure	Average
0.00	0.06	0.06	0.06

Table 22 shows the affinity purification of the antibodies to the lung antigen, L552S.

5

Table 22

Antibody dilution	Affinity pure (salt peak)	Affinity pure (salt peak)	Average	Antibody dilution	Affinity pure (acid peak)	Affinity pure (acid peak)	Average
1:50	0.19	0.18	0.18	1:1000	2.22	2.20	2.21
1:100	0.10	0.09	0.10	1:2000	2.17	2.10	2.13
1:200	0.06	0.06	0.06	1:4000	2.05	2.06	2.06
1:400	0.06	0.06	0.06	1:8000	1.95	1.94	1.95
1:800	0.06	0.05	0.06	1:16000	1.86	1.82	1.84
1:1600	0.06	0.06	0.06	1:32000	1.64	1.57	1.61
1:3200	0.06	0.06	0.06	1:64000	1.28	1.26	1.27
1:6400	0.06	0.06	0.06	1:128000	0.88	0.87	0.88
1:12800	0.06	0.06	0.06	1:256000	0.56	0.56	0.56
1:25600	0.06	0.06	0.06	1:512000	0.35	0.34	0.34
1:51200	0.06	0.06	0.06	1:1024000	0.18	0.17	0.18
0.00	0.06	0.06	0.06	0.00	0.06	0.06	0.06

Tables 23A and 23B show the affinity purification and the Protein A purification of the antibodies to the lung antigen L985P peptide #3482.

10

Table 23A

Purified Antibody Concentration (µg/ml)	Affinity pure	Affinity pure	Average
3.0	1.73	1.69	1.71
1.50	1.32	1.28	1.30
0.75	0.96	0.91	0.93
0.38	0.63	0.61	0.62
0.19	0.38	0.39	0.38
0.09	0.19	0.20	0.19
0.05	0.16	0.10	0.13
0.02	0.12	0.15	0.13

Table 23B

Purified Antibody Concentration (µg/ml)	Pro A pure	Pro A pure	Average
40.70	1.80	1.91	1.85
20.35	1.54	1.53	1.53
10.18	1.04	1.18	1.11
5.09	0.71	0.78	0.74
2.54	0.43	0.52	0.48
1.27	0.25	0.28	0.26
0.64	0.13	0.20	0.17
0.32	0.10		0.10

5

EXAMPLE 21

CLONING OF EXTENDED CDNA SEQUENCE OF L1439P

10

The partial cDNA sequence for clone #61144 (SEQ ID NO:1761, referred to as clone L1439P) was identified from a small cell lung carcinoma PCR based subtraction library on lung chip 5. As previously disclosed in Example 4, microarray analysis was performed on this subset of the lung tumor sequences in order to further evaluate their

expression profiles in various tumor and normal tissues, the results of which are provided in Table 11. Clone L1439P was shown to have a greater than 2-fold over-expression in the tumor probe group versus normal tissues and an average expression in normal tissues of less than 0.2. Additional studies showed this same clone to be over-expressed in a subset
 5 of small cell lung carcinomas (*see*, Table 16).

Further studies have resulted in the isolation of the full-length cDNA for the cloned sequence of L1439P. In order to determine the transcript size of the gene, a multiple tissue Northern blot was probed with the radioactively labeled original cloned sequence (SEQ ID NO:1180). The Northern blot included about 10 ug of total RNA from small cell
 10 lung carcinoma and normal tissues samples. Visual analysis of the exposed film revealed a single transcript of approximately 2.4 kb. The extended sequence of the clone was obtained by screening a small cell lung carcinoma tumor cDNA library with a radioactively labeled probe of the original cloned sequence (SEQ ID NO:1761). Approximately 120,000 clones from the cDNA library were screened and one independent clone containing a
 15 cDNA insert of 1.5 kb was isolated. This extended cDNA sequence is provided in SEQ ID NO:1931, and the deduced amino acid sequence encoded by this extended cDNA sequence is provided in SEQ ID NO:1932. This extended cDNA sequence was analyzed against the Genbank database and was found to show similarity to the NUF2R gene (AF326731). The full-length cDNA sequence of NUF2R is provided in SEQ ID NO:1933, and the deduced
 20 amino acid sequence encoded by NUF2R is provided in SEQ ID NO:1934.

EXAMPLE 22

EXPRESSION IN MEGATERIUM OF A HISTIDINE TAG-FREE L552S FUSION 25 PROTEIN

PCR was performed on the L552S coding region with the following primers:

Forward primer PDM-737 5' ctatgttgcatgcggtgccacgccc 3' (SEQ ID NO:1935) Tm 66°C.

5

10

96°C for 2 minutes, 96°C for 20 seconds, 63°C for 15 seconds, 72°C for 1 minute with 40 cycles and then 72°C for 4 minutes.

15

The PCR product was digested with SphI and BglII restriction enzymes, gel purified and then cloned into pMEG-3, which had been digested with SphI and BglII restriction enzymes. The correct construct was confirmed by DNA sequence analysis and then transformed into *Megaterium* cells for expression.

The amino acid sequence of expressed recombinant L552S is shown in SEQ ID NO:1937, and the DNA coding region sequence is shown in SEQ ID NO:1938.

20

EXAMPLE 23

PCR was performed on the L552S coding region with the following primers:
Forward primer PDM-736 5' ctatgttgcatatatgcggtgccacgcc 3' (SEQ ID NO:1939) Tm 64°C.

25

Reverse primer PDM-480 5' ctgagaattcattaaactgtgggtgctcttcacc 3' (SEQ ID NO:1920) TM 62°C.

The PCR conditions were as follows:

10µl 10X Pfu buffer

1.0μl 10mM dNTPs
 2.0μl 10μM each primer
 83μl sterile water
 1.5μl Pfu DNA polymerase (Stratagene, La Jolla, CA)
 50ng DNA

5

96°C for 2 minutes, 96°C for 20 seconds, 63°C for 15 seconds, 72°C for 1 minute with 40 cycles and then 72°C for 4 minutes.

The PCR product was digested with NdeI and EcoRI restriction enzymes, gel purified and then cloned into pPDM, a modified pET28 vector, which had been
 10 digested with NdeI and EcoRI restriction enzymes. The correct construct was confirmed by DNA sequence analysis and then transformed into BLR pLys S and HMS 174 pLysS cells for expression.

The amino acid sequence of expressed recombinant L552S is shown in SEQ ID NO:1940, and the DNA coding region sequence is shown in SEQ ID NO:1941.

15

EXAMPLE 24

GENERATION OF L552S-SPECIFIC CTL LINES USING IN VITRO WHOLE-GENE PRIMING

20

Using *in vitro* whole-gene priming with tumor antigen-vaccinia infected DC (Yee et al, *The Journal of Immunology*, 157(9):4079-86, 1996), human CTL lines were derived that specifically recognize autologous fibroblasts transduced with the L552S tumor antigen, as determined by interferon-gamma ELISPOT analysis. Specifically, dendritic cells (DC) were differentiated from Percoll-purified monocytes derived from PBMC of
 25 normal human donors by growing for five days in RPMI medium containing 10% human serum, 50 ng/ml human GM-CSF and 30 ng/ml human IL-4. Following culture, DC were infected overnight with a recombinant adenovirus that expresses L552S at a multiplicity of infection (M.O.I) of five, and matured overnight by the addition of 2 μg/ml CD40 ligand. Virus was then inactivated by UV irradiation. CD8+ cells were enriched for by the

depletion of CD4+ and CD14+ cells. CD8+ T cells were isolated using a magnetic bead system, and priming cultures were initiated in individual wells of six 96-well plates with the cytokines IL-6 and IL-12. Cultures were restimulated every 7-10 days using autologous primary fibroblasts retrovirally transduced with L552S, and the costimulatory molecule CD80 in the presence of IL-2. Following three stimulation cycles, two CD8+ T cell lines, 2-12G and 4-7A, were identified using interferon-gamma ELISPOT analysis that specifically produce interferon-gamma when stimulated with the L552S tumor antigen-transduced autologous fibroblasts, but not with a control antigen. Both lines were restimulated and tested again in an antibody blocking assay to determine restriction to specific HLAs. Line 2-12G appears to be HLA-B/C restricted, while Line 4-7A appears to be HLA-A restricted. Line 2-12G was cloned using anti-CD3 and feeder cells, with fourteen specific clones being recovered. These clones have the same pattern of reactivity in antibody blocking assays as the parental L2-12G CTL line. In addition, using a panel of HLA-mismatched B-LCL lines transduced with a vector expressing L552S, and measuring interferon-gamma production by the CTL lines in an ELISPOT assay, these CTLs appear to be restricted by HLA-B*4402.

EXAMPLE 25

20 EPI TOPE-MAPPING OF L552S- AND XAGE-1-SPECIFIC ANTIBODIES

It was previously found that L552S is an alternative splicing isoform of XAGE-1. In this example, data was obtained using 20mer peptides specific for either L552S or XAGE-1 to screen the sera of lung cancer patients for antibodies specific for L552S and XAGE-1, respectively. It was found that individual cancer patients produce both antibodies specific for L552S as well as for XAGE-1. It was determined that these specific antibodies recognize primarily the following additional epitope of L552S and two epitopes of XAGE-1:

L552S-specific:

aa31-50: LGCCWGYPSRSTWDRPF (SEQ ID NO:1942)

XAGE-1-specific:

aa11-30: CSLGVFPSAPSPVWGTRRSC (SED ID NO:1943)

5 aa41-50: ILSPLLRHGGHTQTQNHTAS (SED ID NO:1944).

EXAMPLE 26

IMMUNOHISTOCHEMISTRY ANALYSIS OF L552S

10 In order to determine L552S protein expression in various normal and lung cancer tissues, immunohistochemistry (IHC) analysis was performed using an affinity purified L552S polyclonal antibody. Specifically, tissue samples were fixed in a formalin solution for 12-24 hrs and embedded in paraffin before being sliced into 8 micron sections. Steam heat induced epitope retrieval (SHIER) in 0.1 M sodium citrate buffer (pH 6.0) was
15 used for optimal staining conditions. Sections were incubated with 10% serum/PBS for 5 minutes. Primary antibody was added to each section for 25 minutes at indicated concentrations followed by a 25 minute incubation with either anti-rabbit or anti-mouse biotinylated antibody. Endogenous peroxidase activity was blocked by three-1.5 minute incubations with hydrogen peroxidase. The avidin biotin complex/horse radish peroxidase
20 (ABC/HRP) system was used along with DAB chromogen to visualize L552S expression. Slides were counterstained with hematoxylin to visualize cell nuclei.

IHC analysis of L552S is disclosed in Table 24:

Table 24

Tissue Type	Staining (pos/total)	Comments
Adeno Lung Cancer	10/12	Strong staining (nuclear “dot-like”)
Squamous Lung Cancer	4/12	Light staining (cytoplasmic “dot-like”)
Adrenal	0/1	
Artery-endothelium	0/1	Light cytoplasmic staining
Blood (bone marrow)	0/1	
Brain (cerebellum)	0/1	
Brain (cortex)	0/1	
Breast	0/1	
Bronchus	0/5	
Colon	1/4	Very light staining (cytoplasmic “dot-like”)
Esophagus	0/1	
Fallopian Tube	0/1	
Gall Bladder	0/1	
Heart	0/1	
Kidney	2/4	Very light staining (cytoplasmic “dot-like”)
Liver	1/4	Very light staining (cytoplasmic “dot-like”)
Lung	0/13	
Pancreas	0/1	
Pituitary	1/1	Very light staining (cytoplasmic “dot-like”)
Placenta	1/1	Very light staining (cytoplasmic “dot-like”)
Prostate	0/1	
Skeletal Muscle	0/1	
Skin	0/1	
Small Bowel	0/1	
Spinal Cord	0/1	
Spleen	0/1	
Stomach	0/1	
Testis	1/2	Very few selected cells positive
Thymus	0/1	
Thyroid	0/1	
Trachea	0/1	
Urinary Bladder	0/1	
Ureter	0/1	
Uterus	0/1	

EXAMPLE 27IDENTIFICATION OF CD4 IMMUNOGENIC T CELL EPITOPES
DERIVED FROM L552S

5

CD4 T cell lines specific for the antigen L552S (SEQ ID NO:809) were generated as follows. A total of thirty 20-mer peptides overlapping by 15 amino acids corresponding to the amino acid residues of full-length L552S (SEQ ID NO:809) were synthesized. The amino acid sequence of each of the thirty 20-mer peptides and the
 10 respective DNA sequence which encodes each of these peptides is provided in Table 25. Dendritic cells (DC) were differentiated from Percoll-purified monocytes derived from PBMC of normal male human donors by plastic adherence and growing for five days in RPMI medium containing 10% human serum, 50 ng/ml human GM-CSF and 30 ng/ml human IL-4. Purified CD4 T cells were generated from the same donor as the DCs by
 15 using MACs beads and negative selection of PBMCs. The DCs were pulsed overnight with pools of the 20-mer peptides with each peptide at an individual concentration of 0.5 µg/mL. The pulsed DCs were washed and plated at 10,000 cells per well of 96-well round bottom plates, and purified CD4 T cells were added at 100,000 cells per well. Cultures were supplemented with 10 ng/mL IL-6 and 5 ng/mL IL-12 and incubated at 37°C

20 Cultures were restimulated as above on a weekly basis using DCs made and pulsed as above as the APC, supplemented with 10 U/mL IL-2 and 5 ng/mL IL-7. Following three in vitro stimulation cycles (the initial priming + two restimulations), lines (each line corresponds to one well) were tested for specific proliferation and cytokine production in response to the stimulating pool versus an irrelevant peptide pool of peptides
 25 derived from assorted unrelated antigens. A number of individual CD4 T cell lines (49/576 by IFN-gamma and 63/576 by proliferation) demonstrated significant cytokine release (IFN-gamma) and proliferation in response to the L552S peptide pools, but not to the control peptide pool. Twenty five of the T cell lines which exhibited specific activity were restimulated on the appropriate pool of L552S peptides and reassayed on autologous DCs

pulsed with the individual peptides or recombinant protein made in *E. coli*. Approximately 13 immunogenic peptides were recognized by the T cells from the entire set of peptide antigens tested. These 13 peptides are (*) in Table 25.

In some cases the peptide reactivity of the T cell line could be mapped to a single peptide but some could be mapped to more than one peptide in each pool. This result indicates that all 13 peptides may be naturally processed epitopes of the L552S protein.

Table 25

10

L552S CD4 Overlapping Peptides			
Peptide	Sequence	Peptide SEQ ID NO:	DNA SEQ ID NO:
1	MRCHAHGPSCLVTAITREEG	1945	1974
2	HGPSCLVTAITREEGGPRSG	1946	1975
3	LVTAITREEGGPRSGGAQAK	1947	1976
4*	TREEGGPRSGGAQAKLGCCW	1948	1977
5*	GPRSGGAQAKLGCCWGYPSP	1949	1978
6*	GAQAKLGCCWGYPSPRSTWN	1950	1979
7*	LGCCWGYPSPRSTWNPDRRF	1951	1980
8	GYPSPRSTWNPDRRFWTPQT	1952	1981
9	RSTWNPDRRFWTPQTGPGE	1953	1982
10	PDRRFWTPQTGPGEGRHERH	1954	1983
11	WTPQTGPGEGRHERHTQTQN	1955	1984
12	GPGEGRHERHTQTQNHTASP	1956	1985
13	RHERHTQTQNHTASPRSPVM	1957	1986
14	TQTQNHTASPRSPVMESPKK	1958	1987
15*	HTASPRSPVMESPKKKNQQL	1959	1988

16	RSPVMESP KKKNQQLKVGIL	1960	1989
17	ESPKKKNQQLKVGILHLGSR	1961	1990
18*	KNQQLKVGILHLGSRQKKIR	1962	1991
19*	KVGILHLGSRQKKIRIQLRS	1963	1992
20*	HLGSRQKKIRIQLRSQCATW	1964	1993
21*	RQKKIRIQLRSQCATWKVICK	1965	1994
22*	IQLRSQCATWKVICKSCISQ	1966	1995
23*	SQCATWKVICKSCISQTPGIN	1967	1996
24*	KVICKSCISQTPGINLDLGS	1968	1997
25	SCISQTPGINLDLGSGVKVK	1969	1998
26	TPGINLDLGSGVKVKIIPKE	1970	1999
27*	LDLGSGVKVKIIPKEEHCKM	1971	2000
28	GVKVKIIPKEEHCKMPEAGE	1972	2001
29	IIPKEEHCKMPEAGEEQPV	1973	2002

EXAMPLE 28

HLA RESTRICTION OF L552S

5

To determine the HLA restriction, a panel of fibroblasts matched at one or two HLA alleles were transduced with pBIB L552 or infected with adenovirus L552 at an MOI of 50:1. These targets were tested against a D77 L552 CD8 clone in an ELISPOT assay with 10000 fibroblasts, 10000 T cells and 5 U/mL IL-2 per well. The results indicate

10 that the restriction is with either B*4402 or Cw*0501.

To decide between the two, COS-7 cells were transfected with a combination of pcDNA3 L552S and pcDNA3 B*4402 or pBIB Cw*0501. These targets were again tested against the same D77 L552 CD8 clone as above in an ELISPOT assay

with 10000 COS-7, 10000 clone 14 and 5 U/mL IL-2 per well. The clone recognized the L552 and B*4402 transfected COS-7 cells.

In further studies, three different tumors, 659-22, 3-90T and HTB 183, that had been previously analyzed by real time RT-PCR for L552 message level were selected and transduced with B*4402 or Cw*0501 as a control. 659-22 and 3-90T contain L552 message, while HTB 183 was the negative control. After two rounds of selection, the tumors were analyzed by FACs for their B44 expression level. 659-22 were found to endogenously express high levels of B44, but it was not determined which of the B44s, B*4402, B*4404, etc, were being expressed. Approximately 20% of the 3-90T express B44. The HTB 183 express B44 quite well. When the above clone was tested against these tumors, 659-22 transduced with B*4402 was recognized. Thus, the results indicate that L552 is restriction by HLA-B*4402.

15

EXAMPLE 29

EXPRESSION OF L552S AND XAGE-1

In this Example, real-time RT-PCR analysis was performed in order to delineate the expression of L552S from XAGE-1 in lung cancers. The real-time RT-PCR analysis was performed using specific primers localized in the 5' unique region of L552S, the 5' unique region of XAGE-1 and the 3' common region. Specific messages for L552S and XAGE-1 were detected in two lung tumor samples but not in the normal lung samples.

Identical expression profiles were observed between 5' unique region of L552S and the common 3' sequences. The message level for XAGE-1 detected in lung tumors using the 5' unique primers of XAGE-1 was much lower compared with L552S. However, the extreme secondary structure posed by XAGE-1 could hamper the cDNA synthesis of the 5' sequence unique to XAGE-1. Thus, it appears from the Northern analysis that XAGE-1 may be a more abundant isoform.

EXAMPLE 30

EST EXPRESSION PROFILE OF L552S AND XAGE-1

5 To further evaluate the expression profile of L552S and XAGE-1, an electronic express profiling was performed for each antigen. This was done by searching with the same specific primers as disclosed in Example 29 against a public EST database. Results of this profiling confirm that there are two isoforms of the gene. The ratio of expression between L552S and XAGE-1 is about 1:5. In addition, L552S and XAGE-1
10 seem to be expressed in Hepatoma, CML, germ cell tumor, Ewing's Sarcoma, and Alveolar Rhabdomyosarcoma.

EXAMPLE 31

15 L985P LUNG TUMOR ANTIGEN SURFACE EXPRESSION

 Small cell lung carcinoma (SCLC) cell lines, NCI-H69, NCI-H128, HTB-171, HTB-173, HTB-175, and DMS 79, were grown in DMEM (Gibco) containing 10% FBS (Hyclone) at 37° C with 7% CO₂. These growing SCLC cell lines were then
20 subjected to FACS analysis to determine whether L985P is expressed on the surface of these SCLC cell lines using an anti L985P peptide polyclonal sera raised against the predicted excellular region of L985P.

 For FACS analysis, cells were collected and washed with ice cold staining buffer (PBS+1%BSA +Azide + 10 µg/ml human IgG). The cells were next incubated for
25 one hour on ice either with no primary antibody, with irrelevant rabbit IgG, whole molecule at a final concentration of 20 µg/ml, or with affinity purified rabbit polyclonal sera raised against L985P peptide #3482 (SEQ ID NO:1930) at a final concentration of 20 µg/ml. The sequence of the L985P peptide #3482 (SEQ ID NO:1930) represents the predicted extracellular region of the L985P protein. The cells were washed 2 times with staining

buffer and then incubated with a 1:100 dilution of a goat anti-rabbit Ig(H+L)-FITC reagent (Southern Biotechnology) for 30 minutes on ice. Following two washes, the cells were resuspended in staining buffer containing propidium iodide (PI), a vital stain that allows for identification of permeable cells, and analyzed by FACS.

- 5 In addition, Real-time PCR was performed to determined if L985P mRNA was expressed in these SCLC cell lines. The results of the FACS analysis and the Real time PCR of mRNA expression are presented in Table 26:

Table 26

10

SCLC Cell Line	mRNA Expression by Real Time PCR	Surface Expression by FACS
NCI-H69	+	+
HTB-173	+	+
HTB-175	-	-
HTB-171	-	-
NCI-H128	-	-
DMS 79	-	-

- 15 From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.